Ana Rita Castro Otrelo Cardoso

Mestre em Biotecnologia

Structural studies on molybdenum-dependent enzymes: from transporters to enzymes

Dissertação para obtenção do Grau de Doutor em Bioquímica
Especialidade Bioquímica Estrutural

Orientador: Doutora Teresa Santos Silva
Investigadora Auxiliar
Faculdade de Ciências e Tecnologia - UNL

Co-orientador: Doutora Maria João Romão
Professora Catedrática
Faculdade de Ciências e Tecnologia - UNL

Júri

Presidente: Doutora Maria Luísa Dias de Carvalho de Sousa Leonardo

Arguentes: Doutora Sandra de Macedo Ribeiro
Doutora Inês Antunes Cardoso Pereira

Vogais: Doutor Carlos Alberto Gomes Salgueiro
Doutora Manuela Alexandra de Abreu Serra Marques Pereira

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Dezembro 2017

Universidade Nova de Lisboa
Faculdade de Ciências e Tecnologia

Structural studies on molybdenum-dependent enzymes:

from transporters to enzymes

Ana Rita Castro Otrelo Cardoso

13 December 2017

“**Structural studies on molybdenum-dependent enzymes: from transporters to enzymes**”

O trabalho apresentado nesta Tese foi realizado no âmbito da Bolsa de Doutoramento Individual SFRH/BD/85806/2012 e dos projetos PTDC/BIA-PRO/118377/2010 e PTDC/BBB-BEP/1185/2014 financiados pela Fundação para a Ciência e a Tecnologia - Ministério da Ciência, Tecnologia e Ensino Superior.

Do trabalho desenvolvido resultaram as seguintes publicações:

1. **Otrelo-Cardoso AR**, Nair RR, Correia MA, Cordeiro RSC, Panjkovich A, Svergun DI, Santos-Silva T, Rivas MG. Highly selective tungsten transporter TupA protein from *Desulfovibrio alaskensis* G20. Sci Rep **2017**; 7(1): 5798. DOI:10.1038/s41598-017-06133-y

2. Correia MA*****, **Otrelo-Cardoso AR***, Schwuchow V, Clauss KGVS, Haumann M, Romão MJ, Leimkühler S, Santos-Silva T. The *Escherichia coli* periplasmic aldehyde oxidoreductase is an exceptional member of the xanthine oxidase family of molybdoenzymes. ACS Chem Biol **2016**; 11(10): 2923–35. DOI: 10.1021/acschembio.6b00572. *****These authors contributed equally to this work.

3. **Otrelo-Cardoso AR**, Nair RR, Correia MA, Rivas MG, Santos-Silva T. TupA: a tungstate binding protein in the periplasm of *Desulfovibrio alaskensis G20*. Int J Mol Sci **2014**; 15(7): 11783-98. DOI: 10.3390/ijms150711783

4. **Otrelo-Cardoso AR**, Schwuchow V, Rodrigues D, Cabrita EJ, Leimkühler S, Romão MJ, Santos-Silva T. Biochemical, stabilization and crystallization studies on a molecular chaperone (PaoD) involved in the maturation of molybdoenzymes. PlosOne **2014**; 9(1): e87295. DOI: 10.1371/journal.pone.0087295

5. **Otrelo-Cardoso AR**, Correia MA, Schwuchow V, Svergun DI, Romão MJ, Leimkühler S, Santos-Silva T. Structural data on the periplasmic aldehyde oxidoreductase PaoABC from *Escherichia coli*: SAXS and preliminary X-ray crystallography analysis. Int J Mol Sci **2014**; 15(2): 2223-36. DOI: 10.3390/ijms15022223

Agradecimentos

Esta tese é dedicada ao meu marido e companheiro de aventuras, Mílton Cordeiro. Estás sempre ao meu lado e continuas a dar-me a mão e a levar-me cada vez mais longe. Sem ti nunca teria conseguido chegar aqui. Esta tese também é tua. Obrigada por tudo! Pelo amor, pelo apoio, por me fazeres feliz todos os dias, por me teres dado o nosso bem mais precioso... O melhor ainda está para vir!

Ao meu filho, Vasco. És a maior alegria da minha vida! O meu maior e melhor projecto! És um amor que não se explica, que nos faz ser melhores e que nos ensina a valorizar o que é realmente importante! Ver-te crescer é a melhor recompensa que podemos ter.

À minha querida mamã, Anabela, porque te devo tudo aquilo que sou! Obrigada por acreditares e confiares em mim! Obrigada por termos deixado tudo para trás para sermos felizes!

Às minhas orientadoras, Doutora Teresa Santos-Silva e Professora Maria João Romão, um profundo e sincero obrigada pela oportunidade e por todo o apoio prestado ao longo destes anos. Foram anos de muita aprendizagem, de enriquecimento profissional e pessoal.

À Márcia Correia e Raquel Cordeiro que foram peças fundamentais para o trabalho aqui apresentado. Nunca vou esquecer os momentos e as risadas que demos juntas!

À Professora Silke Leimkülher, Viola Schwuchow, Nadine Böhmer e Maria Gabriela Rivas pela colaboração.

Aos meus avós Luz e João, tios Jorge e Lurdes e primo Francisco, que me viram crescer e muito contribuíram para eu ser o que sou.

Aos meus queridos sogros, Eugénia e José António, pelo carinho e suporte.

Aos meus amigos, que me enchem o coração e confortam a alma: Sara, Susana, Mónica, Catarina G, Diana R, Pedro, André, Chagas, Joanas, Saúl… e todos os outros que trago no coração.

Ao Filipe, à Catarina e ao Francisco, por aturarem e apaziguarem o meu mau humor matinal e pela amizade para lá da bancada. Até que os Dim Sum nos separem…

Aos meus colegas: Marino, Viviana, Benedita, Jorge, Cecília, Raquel C., Ana Luísa e Angelina, por todo o apoio, por tudo o que me ensinaram e pelos bons momentos.

Ao Hugo.

Aos sempre presentes e igualmente importantes, Snoopy e Gibbs!

Muito obrigada a todos!

'Yesterday is history, tomorrow is a mystery, today is a gift (…)'

Bil Keane

Abstract

Molybdenum (Mo) and tungsten (W) are heavy metals that can be found in the active site of several enzymes important for the metabolism of carbon, sulfur and nitrogen compounds. This Thesis describes the structural studies of two proteins that are involved in Mo and W uptake (TupA and ModA), of a Mo-containing aldehyde oxidoreductase (PaoABC) and of its chaperone PaoD. The main techniques used for the structural characterization of these proteins are X-ray crystallography and Small-Angle X-ray Scattering (SAXS), which are presented in Chapter 1, including a brief introduction about the importance of Mo and W in biological systems.

Mo or W cofactor biosynthesis requires the presence of molybdate and tungstate inside the cells, which is achieved by specific ABC transport systems. Chapter 2 presents a small introduction about these transport systems, followed by the structural characterization and analysis of ModA and TupA from *Desulfovibrio alaskensis G20*. The tridimensional structures were determined by X-ray crystallography and SAXS, and the implication in the molybdate/tungstate uptake and discrimination between ligands discussed. The results show that TupA has a high selectivity for tungstate, while ModA is not able to distinguish between the two oxyanions. An important residue for TupA selectivity was identified, R118, paving the way for future biotechnological applications.

Chapter 3 focuses on Mo-containing enzymes and cofactor maturation. The tridimensional structure of the *Escherichia coli* periplasmic aldehyde oxidoreductase PaoABC was solved at 1.7 Å resolution, revealing the presence of an unexpected [4Fe-4S] cluster that was not previously reported. The PaoABC structure has unique features, being the first example of an heterotrimer (αβγ) from the xanthine oxidase family. The activation of PaoABC is dependent on its interaction with the chaperone PaoD, which was also studied. The stabilization of *E. coli* PaoD is extremely challenging but the results here presented show that the presence of ionic liquids during thawing avoids protein aggregation. This allowed the identification of two promising crystallization conditions using polyethylene glycol and ammonium sulfate as precipitant agents.

Chapter 4 describes the use of SAXS for the characterization of a multi-component biosensor to detect chronic myeloid leukemia, demonstrating the versatility of this technique to determine the envelope of biological molecules as oligonucleotides.

The main conclusions derived from the work here described, as well as future perspectives, are drawn in Chapter 5.

Keywords: X-ray Crystallography • Small-Angle X-ray Scattering • Molybdenum cofactor • Tungsten • ABC transporters • Molybdoenzymes • Chaperones.

Resumo

O molibdénio (Mo) e tungsténio (W) são metais pesados encontrados no centro activo de diversas enzimas que desempenham um papel importante no metabolismo de compostos de carbono, enxofre e azoto. A presente Tese descreve o estudo estrutural de duas proteínas envolvidas no transporte de Mo e W (ModA e TupA) para o interior da célula, uma enzima de molibdénio (PaoABC) e a sua chaperona (PaoD). As principais técnicas utilizadas para esta caracterização estrutural foram Cristalografia de Raios-X e Dispersão de Raios-X de Ângulos Baixos (SAXS), apresentadas no Capítulo 1. Para além da introdução técnica, este capítulo também inclui uma breve introdução sobre a importância do Mo e W em sistemas biológicos.

A síntese dos cofactores de Mo e W requer a presença de molibdato e tungstato no interior das células, sendo esta assegurada por transportadores específicos do tipo ABC. O Capítulo 2 contém uma breve introdução do sistema em causa, e a análise e caracterização estrutural da ModA e TupA de *Desulfovibrio alaskensis* G20. Foram determinadas as estruturas por cristalografia de raios-X e SAXS, e discutidas as implicações na captura e distinção entre ligandos. Os resultados obtidos demonstram que a TupA tem uma maior selectividade para o tungstato, enquanto a ModA liga os oxoaniões de igual forma. Foi identificado um aminoácido importante para a selectividade da TupA (R118), abrindo caminho para futuras aplicações biotecnológicas desta proteína.

O Capítulo 3 centra-se na temática das molibdoenzimas e maturação do cofactor de molibdénio. A estrutura tridimensional da aldeído oxidoredutase periplasmática PaoABC de *Escherichia coli* foi resolvida a 1.7 Å e revelou a existência de um centro [4Fe-4S] que não tinha sido ainda descrito. A estrutura da PaoABC tem características únicas, sendo o primeiro exemplo de um heterotrímero ($\alpha\beta\gamma$) da família da xantina oxidase. A activação desta enzima está dependente da interacção com a sua chaperona PaoD. A presença de líquidos iónicos durante o processo de descongelamento da PaoD aumentou a estabilidade da proteína, o que permitiu a determinação de duas condições de cristalização usando polietilenoglicol e sulfato de amónia como agentes precipitantes.

O Capítulo 4 descreve o uso da técnica de SAXS para a caracterização de um biossensor baseado na tecnologia de *nanobeacons* para a detecção da leucemia mielóide crónica. Esta aplicação demonstrou a versatilidade desta técnica para determinar o envelope de diferentes biomoléculas, nomeadamente oligonucleotídeos.

As principais conclusões derivadas do trabalho aqui descrito, bem com as perspectivas futuras, são apresentadas no Capítulo 5.

Termos chave: Cristalografia de raios-X • Dispersão de raios-X de ângulos baixos • Cofactor de molibdénio • Tungsténio • Transportadores ABC • Molibdoenzimas • Chaperonas.

Table of contents

Figure index

Table index

Abbreviations and symbols

ABC	ATP-binding cassette
AOX	Aldehyde oxidase
BtuCD	ABC importer for vitamin B12 from *Escherichia coli*
***Bt*XO**	Xanthine oxidase from *Bos taurus*
CV	Column volume
CML	Chronic myeloid leukemia
CODH	Carbon monoxide dehydrogenase
cPMP	Cyclic pyranopterin monophosphate
Cryo-EM	Cryo-Electron Microscopy
***Da*G20**	*Desulfovibrio alaskensis G20*
***Dg*AOR**	Aldehyde oxidoreductase from *Desulfovibrio gigas*
DLS	Dynamic Light Scattering
DMSO	Dimethyl sulfoxide
ECF	Energy-coupling factor
ESRF	European Synchrotron Radiation Facility
FAD	Flavin Adenine Dinucleotide
F_{cal}	Calculated structure factor
FDH	Formate dehydrogenase
F_{obs}	Observed structure factor
FRET	Förster Resonance Energy Transfer
GPCR	G-protein coupled receptors
***Hs*AOX1**	Human aldehyde oxidase
HSQC	Heteronuclear Single Quantum Coherence
IL	Ionic liquid
ITC	Isothermal Titration Calorimetry
LB	Luria-Bertani
MAD	Multi-wavelength anomalous diffraction
mARC	Mitochondrial amidoxime reducing component
MB	Molecular beacon
MBP	Maltose-binding protein
MCD	Molybdopterin cytosine dinucleotide
MIC	Microbially influenced corrosion
MIR	Multiple isomorphous replacement
MGD	Molybdopterin guanine dinucleotide
Moco	Molybdenum cofactor
ModA	Molybdate-binding protein
ModABC	Molybdate ABC transporter system
MPT	Molybdopterin

MR	Molecular replacement
NBD	Nucleotide-binding domains
NMR	Nuclear magnetic resonance
NSD	Normalized spatial discrepancy
PaoABC	Periplasmic aldehyde oxidoreductase from *Escherichia coli*
PDB	Protein Data Bank
pI	Isoelectric point
PI	Polydispersity index
RMSD	Root-mean-square deviation
SAD	Single-wavelength anomalous diffraction
SAXS	Small-Angle X-ray Scattering
SBP	Substrate-binding protein
SIR	Single isomorphous replacement
SO	Sulfite oxidase
SDH	Sulfite dehydrogenase
SDS/PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
STD	Saturation transfer difference spectroscopy
***Ta*HBCR**	4-hydroxybenzoyl-CoA reductase from *Thauera aromatica*
TMD	Transmembrane domain
TRAP	Tripartite ATP-independent periplasmic
TTT	Tripartite tricarboxylate transporters
TupA	Tungstate-binding protein
TupABC	Tungstate ABC-transporter system
Woco	Tungsten cofactor
XDH	Xanthine dehydrogenase
XO	Xanthine oxidase

Chapter 1

General introduction

1.1. Molybdenum and tungsten in biological systems

In biological systems, transition metals increase the catalytic diversity that can be achieved when only considering the functional groups of amino acids side chains. Transition metals can coordinate directly to side chains (histidine, serine, cysteine or tyrosine), the backbone carbonyl/amino groups, or be incorporated as part of a larger prosthetic group, and heme-containing proteins are the most famous examples. These structures consist of an iron atom coordinated with a porphyrin ring, with biological functions ranging from oxygen transport to gene expression regulation[1].

Molybdenum (Mo) and tungsten (W) are essential for life and considered as micronutrients: they are essential to maintain cell homeostasis but required in low concentrations. The discovery that Mo and W perform a functional role in biological systems is relatively recent, being reported in 1930 by Bortels *et al*.[1]. In this study, it was demonstrated that Mo acted as a catalyst in the fixation of nitrogen by *Arthorobacter chroococcum*. In 1953, two different research groups found that Mo is crucial for the maintenance of normal levels of the enzyme xanthine oxidase (XO) in rats[2,3]. The evidence that W could also play an important role was demonstrated later, in the early 1970s, with several works from Andreesen *et al.* showing that this metal stimulated the growth of certain *Clostridium* bacteria[4,5].

Although Mo and W are trace elements in the earth's crust (at *ca* 230 and 120 ppm, respectively), they are available to biological systems due to the high solubility of molybdate (MoO_4^{2-}) and tungstate (WO_4^{2-}) oxyanions in water. Nowadays, molybdenum is the most abundant transition metal element in the oceans (~110 nM)[6,7] – Table 1.1.

Table 1.1. The abundance of several elements with biological relevance. Adapted from[7].

Location	Abundance (ppb)						
	Mo	**W**	**Fe**	**H**	**C**	**N**	**O**
Universe	0.1	0.003	20×10^3	930×10^6	500×10^3	90×10^3	800×10^3
Crustal rocks	230	120	23×10^6	31×10^6	3.1×10^3	29×10^3	600×10^6
Ocean	0.64	0.004	0.33	662×10^6	14.4×10^3	220	331×10^6
Human body	7	-	6.7×10^3	620×10^6	120×10^6	12×10^6	240×10^6

Molybdenum and tungsten belong to the sixth group of the periodic table, with the atomic number 42 and 74, respectively. The biological roles of the enzymes containing these metals are fundamental and include the catalysis of key steps in carbon, nitrogen and sulfur metabolism[6,8,9].

Tungsten might have been the first of these two elements to be acquired as a functional element by living organisms. Under anaerobic conditions and high sulfur concentrations known to exist during the origin of life period (which prevail in today's deep-sea hydrothermal vents), tungsten forms relatively soluble salts (as WS_4^{2-}). In this environment, molybdenum occurs as the water-insoluble MoS_2 and thus becomes unavailable for biological systems. It is exactly in these conditions where tungsten-using extremophilic bacteria (archaea) were found[6,7,10]. Besides being found in obligate anaerobic prokaryotes, tungstoenzymes are also found in some aerobic methylotrophic organisms, and one example is the formate dehydrogenase (FDH) from *Methylobacterium extorquens* AM1[11]. Molybdenum is more bioavailable to plants and bacteria since it is present in the soils as MoO_4^{2-} [12]. Both metals are needed in trace and balanced amounts but they are lethal for the organisms at high concentrations. For these reasons, the metals are transported into the cell in the form of the oxyanion (molybdate or tungstate) through a delicately regulated, high-affinity, ATP-binding cassette transporter system (ModABC, WtpABC and TupABC – for bacteria)[13]. Within the cell, Mo/ W are subjected to a complex biosynthetic pathway that ends with the incorporation of the metal in the active site of several enzymes. With exception of the multinuclear $MoFe_7$ cluster present in nitrogenase, molybdenum (and tungsten) is found in all other known Mo(W)-enzymes in a mononuclear form. Here, the metal is coordinated to one/two organic tricyclic pyranopterin cofactor via its dithiolene group, Figure 1.1, that may be present either in the dinucleotide or monophosphate form[14,15]. In eukaryotes, only the monophosphate form (MPT) is present, while in prokaryotes it is often conjugated to nucleosides, usually cytosine (MCD, molybdopterin cytosine dinucleotide) or guanosine (MGD, molybdopterin guanosine dinucleotide), and occasionally adenosine or inosine [15,16] – Figure 1.1.

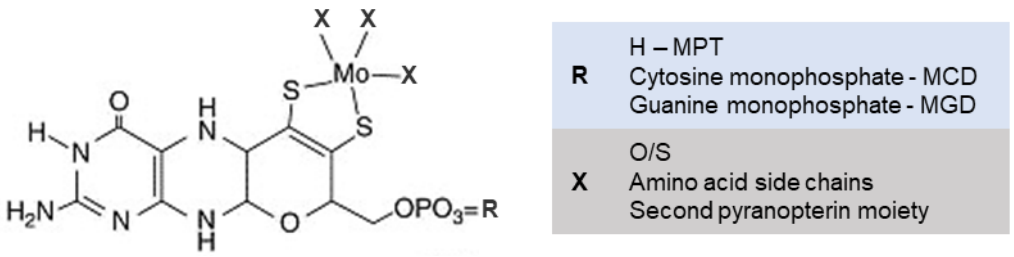


Figure 1.1. The structure of the pyranopterin cofactor present in mononuclear molybdenum and tungsten enzymes. The metal is further coordinated to O/S atoms, and/or amino acid side chains, and/or to a second pyranopterin moiety [17].

The deficiency of the molybdenum cofactor in mammals causes the inactivation of several enzymes that are involved in essential steps, including the catabolism of purines and the metabolism of sulfur-containing amino acids. The molybdoenzymes are also involved in nitrate assimilation, purine metabolism, hormone biosynthesis, and most likely, in sulfite detoxification[18] in plants.

The focus of this Thesis is the study of the selective uptake of tungstate and molybdate by bacterial cells, its incorporation in the active site of important enzymes as cofactors and the structural characterization of a Mo-containing aldehyde oxidoreductase – Figure 1.2. Chapter 2 includes a detailed introduction about the transport of these metals into the cells, while Chapter 3 approaches the molybdenum cofactor biosynthesis and the molybdoenzymes. The next two sections of Chapter 1 contains a brief introduction to the main techniques used to study these pathways: X-ray Crystallography and Small-Angle X-ray Scattering (SAXS).

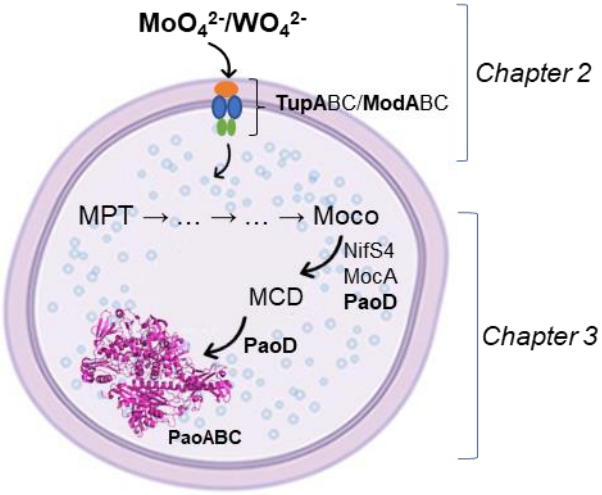


Figure 1.2. Schematic representation of the main topics of the Thesis. The study starts with the uptake of molybdenum or tungsten via specific transport systems, ModABC and TupABC. The metal is the central piece of a biosynthetic pathway that ends with a formation of a Mo/W-cofactor. These cofactors are incorporated in the active site of important enzymes, and the PaoABC is one of the examples.

1.2. Biomolecular crystallography

The 3D structure of a protein is one of the major contributions for its biological characterization and understanding of the biological role. Although other techniques, such as Nuclear Magnetic Resonance (NMR), Cryo-Electron Microscopy (Cryo-EM) and Small-Angle X-ray Scattering (SAXS), have emerged as alternative/complementary techniques, X-ray crystallography is the gold-standard for obtaining atomic resolution information of macromolecules.

The history of biomolecular crystallography starts in the 1950s with John Kendrew and Max Perutz. They determined the first crystal structures of the sperm whale myoglobin[19] and horse hemoglobin[20], respectively. In 1962, they received the Nobel Prize in Chemistry for their studies on globular protein structures. In the same year, James Watson and Francis Crick[21] were awarded the Nobel Prize in Medicine for revealing the double-helix model of DNA, based on the X-ray fiber diffraction, using the images generated by Rosalind Franklin. Two years later, the Nobel Prize in Chemistry was awarded to Dorothy Hodgkin, for her exceptional contributions for solving small molecule structures, such as penicillin, vitamin B12 and cholesterol[22]. These scientists paved the way to the development of biomolecular crystallography, and from the middle last century to nowadays, this field continues to grow with, currently (July 2017), 89.5% of all structures (132055) deposited in the Protein Data Bank (PDB) determined by X-ray crystallography[23]. This technique is used every day to answer important biological questions, with its importance recognized by several Nobel Prizes awarded (from the structure of the DNA to the multi-protein complex, the ribosome) and with 'The International Year of Crystallography' declared by the United Nations in 2014[24].

1.2.1. General concepts

Biomolecular crystallography is based on the interaction of electrons present in the molecules with X-rays. This type of radiation was discovered by the German physicist Wilhelm Röntgen in 1895, and the name resulted from the fact that this was an unknown type of radiation at the time. X-rays are a high-energy electromagnetic radiation with wavelengths ranging between 0.1 and 100 Å, corresponding to the same range of the interatomic distances in molecules (~1.0 Å)[25]. They can be produced in vacuum tubes by bombarding a metal target (usually copper or molybdenum) with electrons, leading to the emission of X-rays with wavelengths dependent on the anode material. The Mo anode generates X-rays with a wavelength of 0.7107 Å, traditionally used for data collection from crystals of small molecules. Macromolecular crystallographers have used in-house sources with Cu anodes with a wavelength of 1.5418 Å and/or synchrotron facilities [26,27].

In the early 20th century, Max von Laue used this powerful discovery and demonstrated that when the X-rays hit a periodic object, as a protein crystal, they are diffracted by the electrons resulting

in a diffraction pattern[28]. The obtained diffraction pattern reflects the composition of the crystal and can be used to calculate an electron density map. From this map, an atomic model can be progressively built and refined. Before the deposition of the atomic coordinates in the PDB, a careful validation is necessary. The different steps involved in the determination of a protein structure are illustrated in Figure 1.3 and will be discussed in detail.

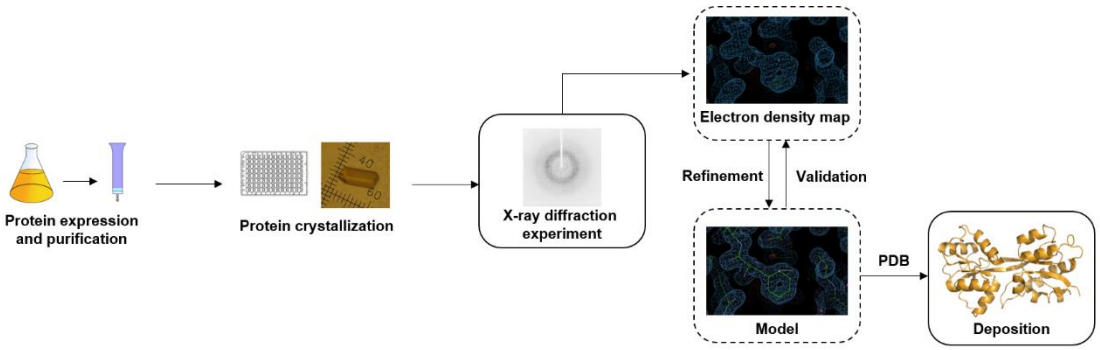


Figure 1.3. Illustration of the most important steps in modern protein X-ray crystallography.

1.2.2. Protein crystals and crystallization

The applicability of X-ray crystallography is dependent on protein crystals, to allow the collection of accurate diffraction intensities. The quality of the final model is directly influenced by the quality of diffraction, so the crystal quality is the key of the entire process and the ultimate determinant of its success. However, the best conditions to obtain a pure stable protein sample may not be the best conditions for crystallization, which complicates the overall process. As the formation of a crystal lattice is a complex process, with multiple variables involved in protein crystallization. Thermofluor[29] or Dynamic Light Scattering[30] (DLS) are routinely used to understand and increase protein stability through the selection of the right buffer composition (pH, additives, salts)[31,32]. Intrinsic protein properties, such as the isoelectric point (pI), are also relevant. For example, in 2015 Kirkwood *et al.* [33] analyzed the X-ray structures deposited in PDB and showed that acidic proteins (pI <7) tend to crystallize at 1.0 pH unit above their pI and basic proteins at 1.5-3.0 pH units below their isoelectric point, supporting the previous study from Kantardjieff *et al*[34] ten years before.

The most common technique for growing a protein crystal is called *vapor diffusion* (Figure 1.4) and can be performed using the hanging (Figure 1.4. A) or sitting drop methods (Figure 1.4. B). A drop containing a mixture of a precipitating agent (such as a salt or a polymer) and protein solution is placed in a sealed chamber with the precipitant solution. Since the concentration of the precipitant is different in the drop and in the reservoir, the water vapor diffuses out of the drop until the osmolarity of the drop and the precipitant equalize. The dehydration causes a slow increase of concentration of both protein and precipitant until equilibrium is achieved, ideally

placing the protein in the crystal nucleation zone of the phase diagram[35] - Figure 1.5. Typically, the protein crystallization process is divided into two steps: nucleation and crystal growth [28,36,37]. These steps require the presence of a supersaturated state (where the protein concentration exceeds the solubility) that acts as a driving force of the crystallization process. In the 'labile' zone occurs nucleation, which is the most difficult state to address since it represents a first-order phase transition by which the protein molecules pass from a wholly disordered state to an ordered one. Here, the supersaturation is large enough to spontaneously form small microscopic clusters of protein – nucleus - from which the crystal will eventually grow[38,39]. The growing and stabilization of crystals occur in the 'metastable' zone, mainly by the classical mechanism of dislocation and growth by two-dimensional nucleation. In this region, no nucleation takes place [36,40]. In the undersaturated zone, the protein is totally dissolved and will not crystallize. Contrarily, in the high-supersaturated region, also known as precipitation zone, protein aggregates and precipitates form faster than crystals [39,41].

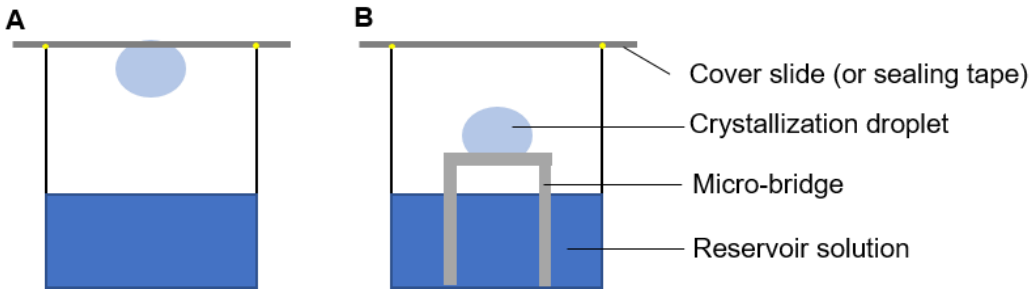


Figure 1.4. Illustration of the vapor diffusion technique using the hanging-drop (A) and sitting-drop (B) methods. In both cases, the drop contains 0.1–10 µl of a protein + precipitant solution mixture. The precipitant is usually the same in the reservoir and in the drop. The water evaporation leads to the equalization of osmolarity of the drop to that of the reservoir, with an increase in the protein and precipitant concentration in the drop.

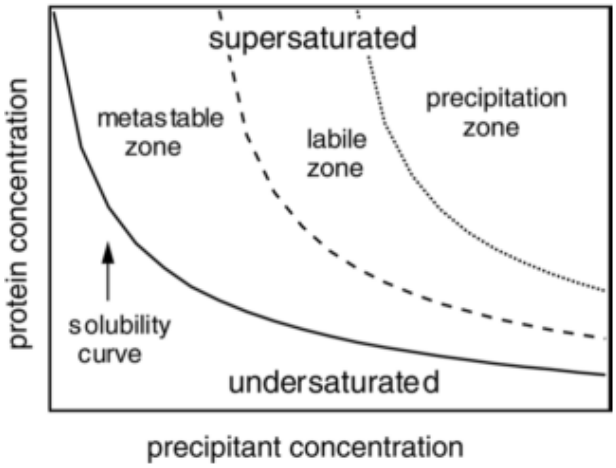


Figure 1.5. Phase diagram for protein crystallization. The diagram contains a region of undersaturation and supersaturation divided by the line denoting the maximum protein solubility at precipitant concentration. The supersaturated region is divided in the metastable zone, where nuclei will grow into crystals, the labile zone (or nucleation zone) and the precipitation zone. Crystals can only grow from a supersaturated solution. Adapted from[39].

The sitting/hanging drop approaches may be the easiest for screening a wide range of crystallization conditions and to get an initial crystallization condition, however are not the best means for optimization. Thus, the vapor diffusion is the elected method to start but ultimately it may be interesting to try another approach better suited for the growth of larger crystals of higher quality. Other alternatives are the micro-batch under-oil and the counter diffusion methods[42,43]. Micro-batch is an alternative when the mother-liquor components cannot be transported through the vapor phase (e.g. metal ions and detergents). The counter diffusion allows testing a wide range of concentrations using one single crystallization assay, which can be recommended for some cases. It also allows *in situ* X-ray data collection at room and cryogenic temperatures and has been employed to grow crystals in microgravity conditions [38,42,44].

The protein crystallization is often a time-consuming step due to the multiple variables that influence the process. The crystallization robots for automated crystallization increase the number of conditions for testing, using a smaller amount of protein, when compared with the traditional manual drop cast methodologies. Despite the difficulties in scale-up the nanoscale crystallization hits, the robots are the easiest way to test different precipitant conditions, additives, drop proportions, and ligands [31,45,46].

Focusing in the crystallography fundaments, crystals are periodic assemblies of identical objects (small or macromolecules) disposed in the tridimensional space. The crystal can be decomposed in a small repeating unit - unit cell – that generates the entire crystal using only translation operations. The regular spacing of the origin of single unit cells is named crystal lattice. The smallest unit that can generate the whole unit cell, using the crystallographic symmetry operators, is called asymmetric unit. The asymmetric unit can be composed by one or more molecules and, in some cases, only includes a part of a functional unit (e.g. a monomer of a functional dimer). In the case of more than one identical molecules, these can be related by non-crystallographic symmetry (NCS) [28,31,45].

The unit cell is defined by the length of three unique edges a, b and c, and three unique angles between them, α, β and γ – Figure 1.6.

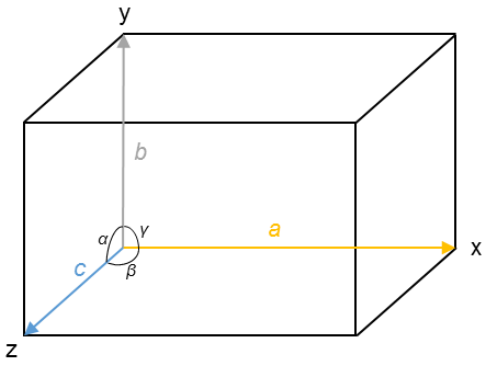


Figure 1.6. Illustration of a unit cell with the angles (α, β, γ) and edges (a, b, c) represented.

Depending on the unit cell constants, seven crystal classes were defined: cubic, tetragonal, orthorhombic, rhombohedral, hexagonal, monoclinic and triclinic. When the crystal classes are combined with the four types of unit cells (primitive (P), face-centered on a single face (C), body-centered (I) and face-centered (F)) leads to the 14 Bravais lattices – Figure 1.7. The symmetry of a unit cell and its contents are described by its space group, which contains information about the internal symmetry between the elements within the cell. 'The International Table of Crystallography, Volume A'[28] compile the different arrangements of the asymmetric units in a cell depending on the 230 space groups available.

The symmetry operations needed to describe unit-cell symmetry are translations, rotations, reflections (mirror plane) and combinations of these like centers of symmetry, screw axes and glide planes. Due to the chirality of the amino acids, mirror planes or inversion centers are allowed but are not found in protein crystals. This limitation on the symmetry of unit cells containing chiral molecules reduces the number of space groups from 230 to 65 [28].

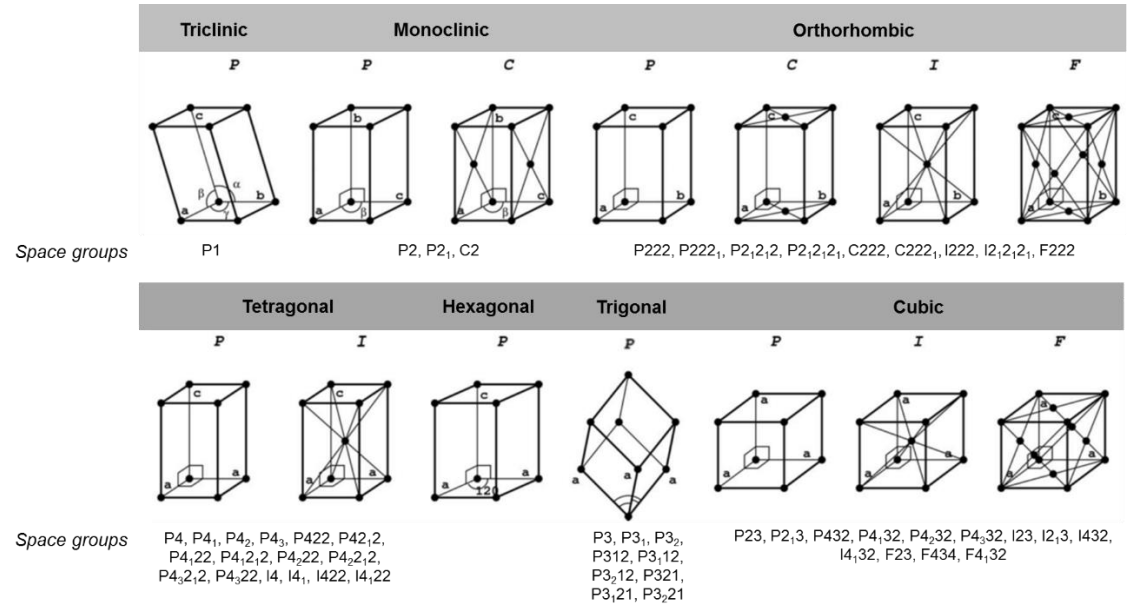


Figure 1.7. The 14 Bravais lattices and space groups allowed in biomolecular crystallography. The black dots represent the lattice points. Types of unit cell: Primitive (P), face-centered on a single face (C), body-centered (I) and face-centered (F). Adapted from [47].

Single protein molecules do not produce a measurable diffraction, hence the need of crystals. The crystal acts as a magnifier of the signal since it contains several ordered copies of the molecule of interest. An ordered crystal packing will diffract the X-ray at high resolution allowing the determination of a correct electron density map. Once obtained a protein crystal, the X-ray diffraction and data collection are the next steps for structure determination.

1.2.3. X-ray diffraction and structure determination

Protein crystals are fragile entities due to the high solvent content, usually in the range of 30-70%[48]. They have large solvent channels, which provide a good access for ligands to bind to protein molecules, through soaking procedures. This physical characteristic leads to the necessity of an extra precaution prior to handling. Usually, the protein crystals need to be pre-equilibrate in a harvesting buffer (which contains a higher precipitant concentration) for stabilization before cryo-cooling (usually, under a cold nitrogen gas, ~100 K) and data collection. Due to the high energy radiation used to obtain the diffraction pattern (especially from a synchrotron source), the data is collected at cryo-temperatures. By minimizing the heat and radiation damage, caused by the formation of free radicals, this procedure allows the collection of a complete dataset [28,49]. To bypass the formation of ice crystals during flash-cooling with liquid nitrogen or cold nitrogen gas, crystals can be soaked in a solution containing a cryoprotectant. Typically, this solution consists in the harvesting buffer supplemented with 20-25% (w/v) glycerol but many other chemical compounds can be used such as sugars, non-detergents or polymers. The formation of crystalline ice can obscure protein diffraction data or even destroy the crystal, compromising the measurement [50,51].

When the X-ray beam hits the crystal, the radiation is scattered by the electrons and results in a diffraction pattern, with reflections on a detector. Each reflection contains information from all atoms in the protein structure [49]. But how the diffraction pattern arises? In 1913, William Lawrence Bragg derived a general equation (Equation 1.1), known as the Bragg's Law, to describe the founding principle of image formation by X-ray diffraction[22,52]. According to Bragg's Law and assuming parallel planes (characterized by the Miller indices (h, k, l)) in the crystal lattice (Figure 1.8), a reflection is collected only when constructive interference of the scattered X-rays occurs [28,53].

$$n\lambda = 2dsin\theta$$

(Equation 1.1.)

In Equation 1.1, n is an integer, λ is the wavelength of the incident radiation, d measures distances in the crystal lattice, also referred to as real lattice, and θ the angle between the incident wave and the scattering planes. The minimum d- spacing corresponds to the highest θ angle at which measurable diffraction has been recorded, known as the resolution of the diffraction pattern [27]. A diffraction pattern is formed only if the difference in the path length of the reflected waves from parallel planes (Figure 1.8) is equal to an integral number of wavelengths ($n\lambda$). If this occurs, the waves are in phase with each other, interfering constructively to produce strong reflections (identified by integer hkl indices). The reflections (or spots) contain the contribution from all the atoms in the crystal at the specific diffraction angle and are recorded by an appropriate detector and stored as a set of reflection intensities $I(hlk)$. Note that these intensities were measured at an angle, θ, dictated by the Bragg's Law (Equation 1.1). The diffraction pattern is defined in a

different space that the crystals, called reciprocal space [28,54,55]. This is so, because the diffraction pattern represents the Fourier transform of the crystal structure, which is in the real space[55] – Equation 1.2.

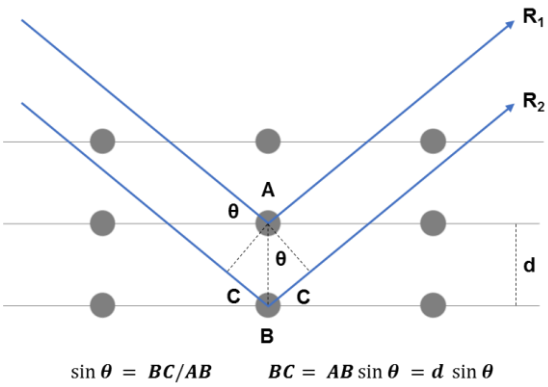


Figure 1.8. Bragg's Law schematic representation. The diffracted X-rays exhibit constructive interference when the distance between paths R_1 and R_2 differ by an integer number (n).

$$F(hkl) = \int_V \rho[cos2\pi(hx + ky + lz) + isin2\pi(hx + ky + lz)]dV$$

(Equation 1.2)

In this equation, the structure factor $F(hlk) = |F(hkl)|e^{i\varphi(hkl)}$ is the wavevector of the corresponding reflection, $I(hkl) = |F(hkl)|^2$. Using the inverse integration of the Fourier transform it is possible to calculate the distribution of the electrons in the unit cell, which corresponds to the electron density, by Equation 1.3[55].

$$\rho(xyz) = \frac{1}{V}\sum_{hkl} F(hkl)[cos2\pi(hx + ky + lz)]$$

or

$$\rho(xyz) = \frac{1}{V}\sum_{hkl}|F(hkl)|e^{i\boldsymbol{\varphi}(hkl)}\, e^{-2\pi i(hx+ky+lz)}$$

(Equation 1.3.)

The data reduction only allows the determination of the moduli $|F(hkl)| = \sqrt{I(hkl)}$ of the structure factors, but not their phases ($\varphi(hkl)$), which are crucial to calculate the electron density map. This limitation is known as the 'Crystallographic phase problem'. Accurate information about the structure factor amplitudes $|F(hkl)|$ is essential for the initial stage of the structure resolution, but also required at the later stages of structure refinement[55,56]. There is no formal relationship

between the amplitudes and their phases. If we have some prior information of the electron density or structure, it is possible to relate them and determine the phases. This is the basis for all phasing methods described in Table 1.2. Following protein crystallization, overcoming the phase problem is the most challenging part of the process.

Table 1.2. Methods for structure solution. Adapted from[56].

Methods	Prior knowledge
Direct methods	$\rho \geq 0$, discrete atoms
Molecular replacement	Structurally similar model
Isomorphous replacement	Heavy-atom substructure
Anomalous scattering	Anomalous-atom substructure

The phases can be determined by direct methods. Here, probabilistic relations between structure factors of certain groups of reflections are used to estimate their phases, usually by expanding a small set of starting phases. This methodology requires diffraction data of, at least, 1.2 Å resolution. They are the methods of choice to determine the structure of small molecules but are not used to solve large macromolecular structures from the native data alone, since the probabilities of phase estimates are inversely proportional to the square-root of the number of atoms [27].

The most common method for solving protein structure is by Molecular Replacement (MR). This method was developed by Rossmann and Blow[57] and can be applied when a structurally similar model is available, usually with a sequence identity of >25%. A Patterson map is calculated using the same Fourier transform described previously for the electron density but using intensities as the coefficients and therefore not requiring the determination of phases. This map has peaks at interatomic vectors rather than at absolute atomic positions. A second Paterson map is determined using the amplitudes calculated from the atomic coordinates (x, y, z) of the search model. From the rotation of search model Patterson map over the Patterson map calculated from the structure-factor amplitudes, the orientation of the model in the new unit cell is obtained. Using also Patterson methods and translation, the position of the model to the origin of the new unit cell is corrected though the comparison of structure-factors between the related models[56]. Despite the power of MR, it is important to be aware of the 'model bias', that occurs when the initial model contains large features of the template model and not the real one [58]. The success of this method is related with the growing number of available structure deposited in PDB and it is very important to assure the quality and accuracy of the models before submission and release to the community (more details about validation in section 1.2.4). The outcome of the presented Thesis contributes with two crystal structure deposited and one under refinement.

In the absence of a suitable homology model, there are very well established *ab initio* methods that can be used, such as the Single/Multiple Isomorphous Replacement (SIR/MIR) and Single/Multiple-wavelength Anomalous Dispersion (SAD/MAD). All require the ordered

introduction/native presence of heavy or anomalous scatterers into the protein crystal. The isomorphous replacement is based on the contribution of the added heavy atom (by soaking or co-crystallization) to the structure-factor amplitudes and phases. Data from a native and derivative crystal are measured. The isomorphous difference between the amplitudes of the two datasets can be used to identify the position of the heavy atoms using the Patterson method. Once located, the atomic coordinates (xyz) of the heavy atoms can be refined and used to calculate a more accurate isomorphous difference and estimate the initial phases. For this method, several crystals are usually required to optimize the soaking or co-crystallization procedure and to ensure the isomorphism between the native and derivative crystals. Usually, several datasets need to be collected until the phase problem can be solved unambiguously [27,28,56,59,60].

The advances in the synchrotron X-ray sources and genetic engineering, makes MAD and SAD the most popular *ab initio* phasing methods. With these approaches, only one well-diffracting crystal is sufficient to solve a structure, so crystal nonisomorphism is not a problem. Typically, the native sulfur-containing methionine of the protein sequence is replaced by an L-seleno-methionine using a methionine auxotrophic *E. coli* strain, introducing the anomalously scattering selenium (with an absorption edge at the wavelength of 0.98 Å) [26,27]. In a MAD experiment, X-rays of a particular wavelength are absorbed by the inner electrons of the selenium atom in the crystal and are re-emitted after a certain delay, inducing a phase shift in all of the reflections (anomalous dispersion effect). This effect, measured as very small differences between datasets collected at different wavelengths, allows the calculation of initial approximate phases[45]. Nowadays, SAD is the method of choice for *ab initio* structure determination with 80% of *de novo* structures being determined by this method. Se-SAD is similar to the Se-MAD experiment except that only one dataset is collected near the selenium absorption edge, where the anomalous scattering signal is greatest ($\Delta f''(Se) = 3.85$). Since it is only necessary to collect data at a fix wavelength, it is possible to perform Se-SAD data collection in an in-house X-ray sources of cooper (λ = 1.54 Å; $\Delta f''(Se) = 1.15$)[61], or chromium (λ = 2.29 Å; $\Delta f''(Se) = 2.30$)[26,62]. Native-SAD is other approach for phasing and uses the anomalous scattering signal of sulfur (in case for proteins) or phosphorous (in case of nucleic acids), inherent atoms, as phasing probes [63]. Anomalous scattering also provides a simple method for overcoming 'model bias' by providing marker atoms and validating the identity of anomalous scatterers for refinement [26].

Once the initial phases and the electron density map are obtained, model building and refinement are the next steps to determine the crystal structure.

1.2.4. Refinement and structure validation

The primary result of an X-ray diffraction experiment is an electron density map. The atomic model is built and refined by varying the model parameters to achieve the best agreement between the F_{obs} (observed reflection amplitudes) and F_{cal} (calculated from the model). The quality of the fit is determined by several crystallographic indicators of data precision[25]. The refinement is an

iterative process with manual corrections and automated optimization that improve the phases and the quality of the electron density map. The optimization involves small adjustments in the atomic coordinates (x, y, z) and B_{factor} (or atomic displacement parameter or temperature factor) of each atom. B_{factor} describes the vibration of an atom around a mean position specified by the atomic coordinates. Well-ordered atoms, usually located in the backbone of α-helixes or β-sheets, have low B_{factor} (5 - 20 Å). On the other hand, side chains and loops that tend to be more flexible are often found in poorly defined electron density area, showing higher B_{factor}[51]. An alternative way of describing atomic displacements involves the segmentation of the whole protein structure into rigid fragments and expressing their vibrations in terms of translational, librational and screw (TLS) movements of each group[25,51].

During refinement, the interpretation of the electron density map requires a significant input of human expertise. A degree of subjectivity is inevitable in this process, thus it is important to have statistical parameters to quantify the discrepancy between the experimental structure factors (F_{obs}) and the calculated from the building model (F_{cal}). The residual or crystallographic R_{factor} (usually, expressed in percentage) is the parameter that allows an overall comparison – Equation 1.4. Depending on the resolution, for well-refine structures a R_{factor} < 20% is expected[28].

$$R_{factor} = \frac{\sum ||F_{obs}| - |F_{cal}||}{\sum |F_{obs}|}$$

(Equation 1.4.)

Due to the characteristics of the refinement procedure, it is important to perform a cross-validation to guarantee the quality of the final model. The indicator R_{free} gives an unbiased measure of agreement, preventing the overfitting during the refinement. It measures, at any stage, how well the current model predicts a random set of measured intensities that were not included in the refinement (usually 5-10% of the reflections). The refinement process is guided by the behavior of R_{factor}/R_{free}, that should converge and decrease during the different stages. The divergence of the two values is an indication that the refinement procedure is not correct and should be re-evaluated. In good quality model, the R_{factor}/R_{free} ratio should be around 20% [28,31].

Parameters like R_{factor} and R_{free} describe the global errors present in the model, and do not consider local errors that might be present. The Ramachandran plot is very useful to verify discrete errors and evaluate the correctness of the backbone conformation of the polypeptide chain. The plot represents the torsion angles, phi ($\boldsymbol{\varphi}$) and psi ($\boldsymbol{\psi}$), of each residue of the protein. A correctly folded polypeptide chain should have > 90% of all residues in the most favored regions of the Ramachandran plot [27,28].

Refinement is an infinite process where, upon reaching a threshold, the gain in terms of the fitting parameter is very minute. Tools such as PROCHECK or MolProbity or WHAT_CHECK allows the

validation of the refined structure and determines if the model is ready for deposition in the PDB[64]. The crystallographers share their knowledge with the scientific community providing an atomic point of view of the biological systems. This technique was the key to understand the role of several proteins from the membrane to the incorporation of molybdenum/tungsten into the enzymes. It was also used as a complementary technique for the structural studies in solution using small-angle X-ray scattering (see next chapter for details).

1.3. Small-angle X-ray scattering

1.3.1. General concepts

Small-angle X-ray scattering (SAXS) is a powerful tool to explore biological macromolecules, providing information about the overall structure and structural transitions in solution at a low resolution (1–2 nm)[65]. The history of SAXS starts in 1939 with Guinier studying metal alloys[66]. Twenty years later, Guinier and Fournet, published the first monograph on SAXS where they demonstrated that the information probed by this approach, was not restricted to the size and shapes of particles but also to the internal structure of disordered and partially ordered systems[67]. With the massive technological advances in synchrotron sources and computational methods, SAXS is currently an established characterization technique with many applications, in particular, to study the overall macromolecular shapes of biomolecules, such as proteins or DNA, in solution [65,68].

SAXS is based on the elastic scattering of X-ray photons by macromolecules. When a monochromatic X-ray beam hits the molecules, the electrons present become sources of secondary waves that are scattered in all directions, upon constructive and destructive interferences. In crystallography, the molecules are arranged in a highly-ordered structure, and these secondary waves result in diffraction peaks that can be used to calculate electron density maps and high-resolution structures. In SAXS, these peaks are not observed due to the random distribution of the molecules in solution. The information regarding the orientation of the molecules is lost but the scattering pattern from the small deflection of radiation (2θ between 0.1 and 10° - small angles) provides information on the magnitude of the interatomic distances of the particles in solution, and allows the determination of the overall structure parameters and size and shape of the molecules[69,70] – Figure 1.9.

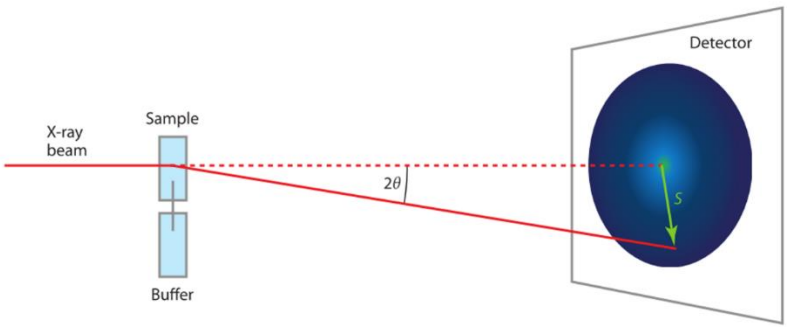


Figure 1.9. A schematic representation of a SAXS experiment. A monochromatic beam hits the solution containing the macromolecules and the scattered photons generate a scattering pattern on a 2D detector. The scattering image is converted to $I(s)$ via radial integration. Adapted from [68].

The X-ray radiation that interacts with the samples is equally scattered in all directions, generating an isotropic scattering pattern. This pattern shows the scattered intensity (I) as a function of the momentum transfer $(s\ or\ q)$ – Equation 1.5.

$$s = \frac{4\pi\, sin\theta}{\lambda}$$

(Equation 1.5)

In Equation 1.5., the θ is half the angle between the incident beam and the scattered radiation and λ is the wavelength of the incident beam[71,72] – Figure 1.9. For a monodisperse solution, the scattering intensity of the biomolecule depends on the concentration and on the contrast between the solute and solvent. The scattering is also influenced by the macromolecule shape and interaction between several particles in solution[70]. When SAXS is applied to biomolecules, the contrast is very small, due to the small difference on the electron density between the solute and solvent. For this reason, SAXS instruments, synchrotron beamlines or in-house sources, must be optimized to minimized the background contribution[73,74].

Considering a dilute monodisperse system, where the biomolecules are in a random position and orientation, the scattering pattern is isotropic, and thus, the scattering collected by a 2D detector can be radially averaged. The background-corrected intensity, $I(s)$, corresponds to the scattering intensity as a function of s (see Equation 1.6) and is proportional to the scattering from a single particle averaged over all orientations (Ω), after subtraction of the solvent scattering [68,75].

$$I(s) = \langle I(s) \rangle_{\Omega} = \langle A(s) A^{*}(s) \rangle_{\Omega}$$

(Equation 1.6)

Here, the scattering amplitude, $A(s)$ – Equation 1.7, is a Fourier transformation of the excess scattering length density (contrast) and $\langle\ \rangle_{\Omega}$ stands for the spherical average.

$$A(s) = \Im[\rho(r)] = \int \Delta\rho(r) \exp(isr)\, dr$$

(Equation 1.7)

In Fourier transformation, $\Delta\rho(r) = \rho(r) - \rho_s$, with $\rho(r)$ and ρ_s corresponding to the electron density of the biomolecule and of the solvent, respectively. These scattering patterns are plotted as radially average 1D curves $I(s)$[76] - example in Figure 1.10. From these curves, several overall important parameters can be directly obtained providing information about the size, oligomeric state and overall shape of the molecule. With the technological advances in X-ray beamlines and computational methods, SAXS also allows for *ab initio* and rigid body modelling, being possible to determine a low-resolution model (1-2 nm) either without any *a priori* information or by using X-ray crystallography or NMR structure as reference[74]. SAXS is also a very useful tool to identify the biologically active conformations of biomolecules in comparison to the crystal structure and clarify oligomeric states. For example, the crystal structure of the Cdt1-Geminin complex was

determined first as a heterotrimer (PDB code 2zxx[77]) and later as a heterohexamer (PDB code 2wvr[78]). From the comparison of the crystallographic data and SAXS data, the authors were able to identified the heterohexamer was the correct model in solution[78].

SAXS can be applied to a broad range of molecular sizes (from a 1 kDa protein to MDa complexes) and requires small amounts of material (typically 1-2 mg protein, 10-100 µL). It is very useful to study the macromolecules in their native conditions but also in the wide range of conditions such as temperature, pH, high pressure, cryo-frozen and chemical or biological additives. Moreover, using a brilliant synchrotron radiation sources it is possible to perform time-resolved experiments that yield unique information about the kinetics of processes and interactions [68,74,76,79].

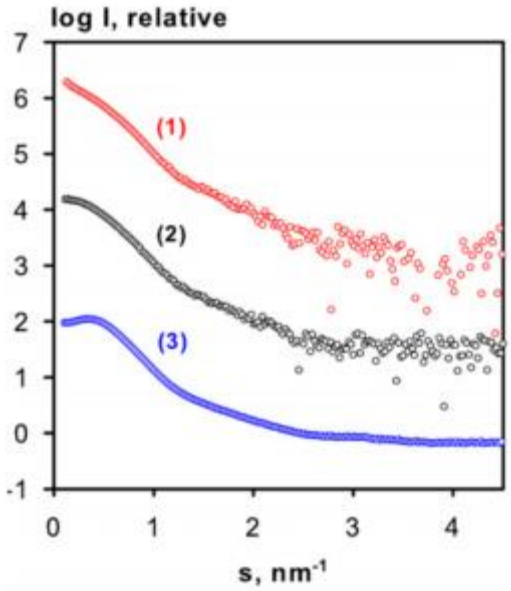


Figure 1.10. SAXS experimental data. Scattering curve of BSA in different buffers showing aggregation (1), good quality data (2) and inter-particle repulsion (3). Adapted from [74].

As previously mentioned, sample scattering intensity is affected by the concentration of the biomolecule and, for this reason, is necessary to measure a range of concentration (e.g. 0.5, 1, 2 and 5 mg/ml). At higher concentrations, the signal-to-noise ratio of the subtracted data is higher, but the distances between the individual molecules are within the same order of magnitude as the intra-particle distances. When a decrease of intensity at low angles is observed, it usually indicates repulsive inter-particles interactions (Figure 1.10 - (3)). In contrast, a sharp increase of intensity points could indicate attractive interactions, which may lead to unspecific aggregation of the sample (Figure 1.10 - (1)). The concentration effect can be minimized by merging the low-angle data at low concentrations with the high-angle data from the higher concentration to yield the final scattering curve. The study of the concentration-dependent behavior of the proteins, for example, can help to define crystallization conditions, which typically require weak attractive interactions[70,76].

By measuring several concentrations it is possible, usually, to eliminate the effect of interactions on the scattering patterns, and extrapolate the scattering curve to infinite dilution that yields the 'ideal' value of the intensity at the zero angle, $I_{ideal}(0)$[76,80,81]. Other important parameters can be obtained directly from the experimental scattering pattern including the radius of gyration (R_g), maximum dimension (D_{max}), molecular weight (MM) and hydrated particle volume (V_P). For a monodisperse solution (ideally higher than 95% of homogeneity), these parameters correspond to the overall characteristics of the molecule. For polydisperse systems, such as intrinsically disorder proteins or aggregates, the values do not correspond to a single molecule, but rather to an average over the entire ensemble[76].

1.3.2. Overall SAXS parameters

The Guinier analysis, developed in 1939, remains the most common and easy method to determine the radius of gyration (R_g) and, consequently the scattering at zero angle $I(0)$. Guinier equation (Equation 1.8) stipulates that, for monodisperse solution and very small angles ($s < 1.3/R_g$), the intensity depends only on two parameters[66,82]:

$$I(s) = I(0)exp\left(-\frac{1}{3}R_g^2 s^2\right)$$

(Equation 1.8)

In practice, R_g and $I(0)$ can be determined by plotting $ln\, I(s)$ vs s^2. The R_g provides information about the mass distribution within the molecule, and is defined as the weighted average of square center-of-mass distances in the molecule. Namely, molecules with the same volume but with different shapes have different R_g values [72,83]. The Guinier plot should be linear, if the measured sample is a pure monodisperse, whereby the slope of the linear region gives R_g and its intersection with the y-axes gives the $I(0)$ – Figure 1.11 (2). A nonlinear plot may suggest an incorrect background subtraction, polydispersity, or inter-particle interactions. In SAXS, it is important to do a prior study of polydispersity since the presence of nonspecific aggregates (Figure 1.11 (1)) or repulsion (Figure 1.11 (3)) between the molecules leads to an overestimation or underestimation of these parameters, respectively [82,83]. The determination of R_g and $I(0)$ is now made automatically by the AUTORG[84] program from ATSAS suite[85].

From the Guinier analysis is possible to determine the molecular weight (MM) of the protein since it is proportional to $I(0)$. This proportionality is determined in the beginning of each data collection through the collection of the scattering data of a standard protein, such as BSA or lysozyme [74,76,86]. This estimation requires normalization against the solute concentrations for the two measurements (protein and standard), and the accuracy of the MM estimate is limited [83,87].

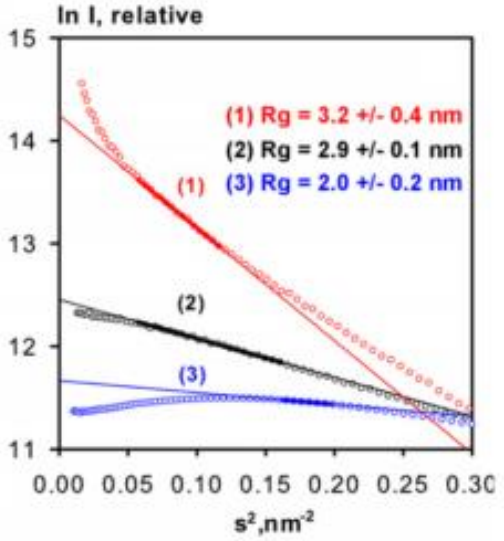


Figure 1.11. Guinier plot of BSA in different buffers showing aggregation (1), good quality data (2) and inter-particle repulsion (3). From [74].

Another important parameter derived from the scattering pattern is the hydrated particle volume (V_p). This parameter is independent of the Guinier analysis, being insensible to the inaccuracies caused by errors in concentration measurements. V_p can be determined by assuming a uniform electron density and using the Porod equation (Equation 1.9), where Q corresponds to the Porod invariant[69].

$$V_P = \frac{2\pi^2 I(0)}{Q}, Q = \int_0^\infty s^2 I(s)\,.ds$$

(Equation 1.9)

To apply this principle to proteins (MM > 30 kDa), an appropriate constant must be subtracted to the scattering profile, generating an approximation of the correspondent homogenous body. Assuming a globular protein, the V_p (in nm^3) can be used to estimate roughly the MM, corresponding to 1.5-2 times of the MM (in kDa)[86].

The R_g and $I(0)$ can be also extracted using indirect Fourier transform methods. Fourier transformation of the scattering intensity yields the distance distribution function, $P(r)$, Equation 1.10:

$$P(r) = \frac{r^2}{2\pi^2}\int_0^\infty s^2\ I(s)\frac{\sin(sr)}{sr}ds$$

(Equation 1.10)

Where the $P(r)$ is real space representation of the distances between all possible pairs of atoms within a molecule and contains information about the shape – Figure 1.12. Due to the limitation on the experimental range of scattering data it is difficult to compute the distance distribution function. This limitation can be overcome by applying an indirect Fourier transformation using the program GNOM[88] (from ATSAS suite), which generates a $P(r)$ from the scattering data base on the D_{max}, is the maximum intraparticle distance[89], defined by the user or by AUTOGNOM[84].

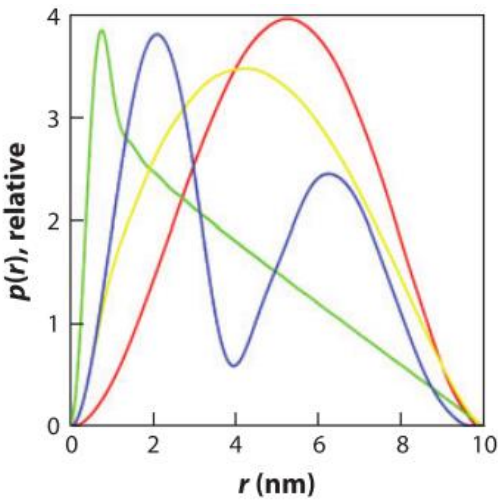


Figure 1.12. Illustration of a distance distribution function for typical geometrical shapes: a sphere (red), dumbbell (blue), cylinder (green) and disk (yellow). From [74].

Usually, a good agreement between the Guinier and real space R_g and $I(0)$ values are an indicator of the dataset quality. The overall parameters can be determined immediately following data collection and are important to characterize the molecules and answer important biological questions.

1.3.3. Molecular shape determination

The determination of a tridimensional shape is important to understand the biological system. The tridimensional models derived from SAXS can be used to complement or can be complemented by other techniques such as X-ray crystallography, NMR or Cryo-EM, being very useful to study protein complexes or different conformations. The molecular envelope is reconstructed via *ab initio* approaches. The determination of the tridimensional shape of molecules derived from the one-dimensional SAXS data started in 90's by Chacón *et al* [90] (in 1998) and Svergun *et al* [73] (in 1999). They developed an *ab initio* method based on automated bead-modeling. The most popular programs for *ab initio* shape reconstruction are DAMMIN (Dummy Atom Model Minimisation)[73], DAMMIF (Dummy Atom Model Minimisation Fast)[91] and GASBOR[92]. They all use simulated annealing to reduce the search space and create an envelope that contains the basic

biomolecules properties. DAMMIN and DAMMIF represent the shape of the biomolecule by densely packed beads with adjustable sizes (typically, a sphere with a diameter equal to the experimentally determined D_{max}). The goal is to minimize the discrepancy (X^2) between the experimental and calculated scattering intensities[93].

GASBOR use dummy atoms, instead of beads, that have the average scattering density of amino acids in water. Here, there is no limitation on the resolution in opposition to the bead model approach, where it is assumed a uniform electron density[86,93]. This program is routinely used to determine the low-resolution structures of proteins and protein complexes[78,92].

One of the major advantages of SAXS is the large size range of biomolecules that can be measured in solution. Large complexes are difficult to study by the most popular methods due to their large dimension, transient nature and flexibility. In some cases, the high-resolution structures of the individual components are available and can be used as a reference (rigid body assembly approach) of the whole complex based on experimental scattering data. Using the program CRYSOL[94], it is possible to calculate the X-ray scattering amplitudes from high-resolution structures and use them as a base for global rigid body modeling. This program uses fast spherical harmonics algorithms to generate SAXS theoretical profiles considering the scattering from the hydration shell [86,94]. The theoretical SAXS curves can be applied to an automated rigid body program, SASREF[95], that performs quaternary structure modeling against single or multiple scattering patterns.

For rigid-body modeling is imperative to have a complete high-resolution model with the coordinates of all components. When domains, loops or purification tags are absent from the reference model, the rigid model cannot be applied directly. The programs BUNCH[95] and CORAL[85] are alternatives that combine the rigid-body and the *ab initio* approaches to model the missing components, as dummy residues.

The cooperation of SAXS with other structural technique is well established and several examples exist in the literature illustrating the multiple applications in different fields, from proteins to nanoparticles. During this Thesis, SAXS was an important tool to clarify the oligomeric state of the periplasmic aldehyde oxidoreductase (PaoABC) from *Escherichia coli*[96] (see Chapter 3) and to study the conformational changes upon ligand binding for two substrate-binding proteins from *Desulfovibrio alaskensis* G20, ModA and TupA (see Chapter 2).

Chapter 2

ATP-binding cassette transporter for tungstate and molybdate in *Desulfovibrio alaskensis* G20

Part of the work described in this chapter was the subject of two publications:

- **Otrelo-Cardoso AR**, Nair RR, Correia MA, Cordeiro RSC, Panjkovich A, Svergun DI, Santos-Silva T, Rivas MG. Highly selective tungsten transporter TupA protein from *Desulfovibrio alaskensis* G20. Sci Rep. **2017**; 7(1): 5798.

- **Otrelo-Cardoso AR**, Nair RR, Correia MA, Rivas MG, Santos-Silva T. TupA: a tungstate binding protein in the periplasm of *Desulfovibrio alaskensis G20*. Int J Mol Sci. **2014**; 15(7): 11783-98.

These two publications are related with the tungstate-binding protein, TupA. The results for the ModA were posteriorly obtained and will be the subject of another publication.

2.1. Introduction

2.1.1. The ABC transporter family

All organism (from humans to a bacteria) rely on the transport of organic and inorganic molecules that cross one or more cell membranes[97]. Cellular survival depends on the passage of specific molecules across these membranes, not only to acquire nutrients and discard waste products but also for regulatory functions. The molecules can pass through the membrane by simple diffusion (typically small and lipophilic molecules), endocytosis/exocytosis (large particles, such as a virus) or by a protein-mediated transport (for large or water-soluble molecules). In the last case, the transport is guaranteed by carrier proteins, or channels that can carry out passive (spontaneous) or active transport (coupled to an energy source) [98]. The importance of membrane transport is evident, with almost ~10% of the *Escherichia coli* genome comprising genes encoding proteins involved in transporting functions, with more than 550 different types of transporters identified[97,99,100]. It is estimated that ~10-60% of the ATP requirements of bacteria and humans (depending on conditions) are used to transport molecules across cell membranes, showing the importance of these proteins to cell homeostasis[97].

ATP-Binding Cassette (ABC) transporters form a superfamily of membrane proteins that are found in all kingdoms of life. Typically, these transporters carry molecules across the lipid bilayers of cellular membranes and convert the energy gained from ATP to ADP hydrolysis into trans-bilayer movement of uptake and efflux of a diverse array of compounds[101–104]. A wide variety of substrates are translocated by this system, from complex molecules such as polysaccharides, peptides and proteins, to smaller components like ions, sugars, amino acids, vitamins, lipids and drugs[105,106]. From a medical perspective, ABC transporters have an enormous interest since they are directly involved in tumor resistance to chemotherapeutics, parasites drug resistance (such *Plasmodium falciparum* or *Leishmania*), fungal drug resistance (like *Candida albicans*), bacterial multidrug resistance, bacterial virulence and pathogenesis (as described for *Streptococcus pneumoniae*)[107–109].

In *E. coli*, the ABC proteins form the largest paralogous group of proteins in this organism[110]. In eukaryotes, ATP hydrolysis occurs in the cytosol, except in mitochondria and chloroplasts where the ATP-binding domains of the transporters are located on the matrix or stroma side, respectively. In prokaryotes, ABC transporters are localized in the plasma membrane with the ATP hydrolysis occurring on the cytoplasmic side. In this context, the termed cis-side and trans-side refer to the side of the cellular membrane where ATP is hydrolyzed or to the opposite side, respectively [104] – Figure 2.1.

ABC transporters can be classified as exporters or importers. ABC exporters are found in prokaryotes and eukaryotes, and transport molecules from the cis-side to the trans-side. In contrast, ABC importers move substrates from the trans-side to the cis-side and seem to be exclusive of prokaryotes organisms[97,104].

All ABC transporters share a basic architecture comprising at least two intracellular nucleotide-binding domains (NBDs) in the cytoplasm and two transmembrane domains (TMDs) – Figure 2.1.

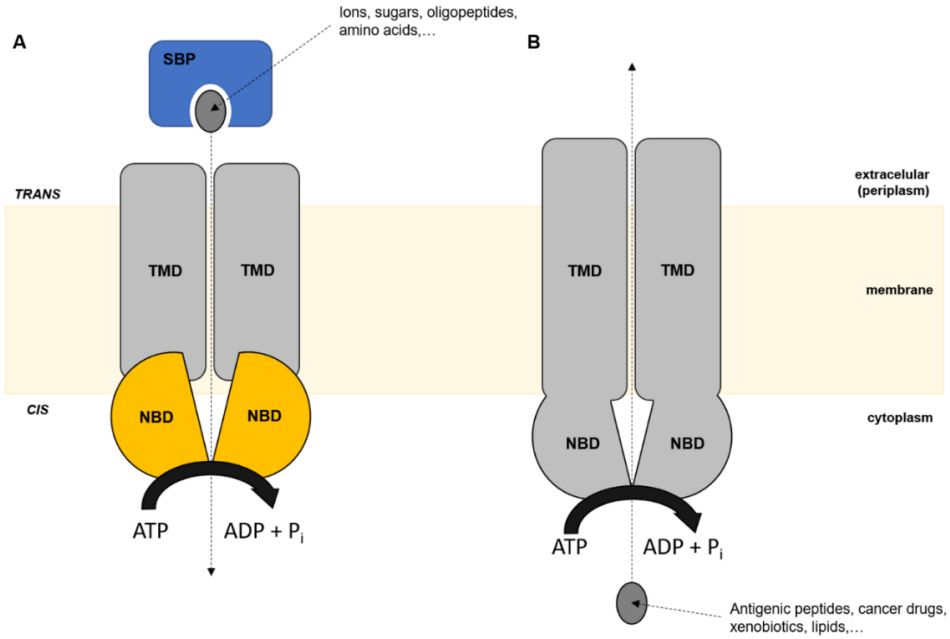


Figure 2.1. Schematic representation of ABC transport system. A) ABC importers. Require a substrate binding protein (SBP) that binds the substrates into the translocation pathway formed by the transmembrane domain (TMD). In this case, the nucleotide-binding domains (NBD) are separate subunits. B) ABC exporters. Typically have their TMDs fused to the NBDs. Adapted from[101].

In prokaryotes importers, besides the TMD and NBD domains, a high-affinity substrate binding protein (SBP) is required, responsible for capturing the substrate and delivering it to the appropriate ABC transporter[104,111,112] (Figure 2.1). In bacteria, the four domains are independent, pairwise identical subunits, or a combination of fused NBDs and/or TMDs. In Gram-negative bacteria, the SBP is a free soluble protein that is present in the periplasm, while in Gram-positive, it is a lipoprotein anchored to the outer leaflet of the plasma membrane. In eukaryotes, most ABC transporters are formed by a single polypeptide that contains all four functional units, with some members assembled as a homodimer or heterodimer[113].

2.1.2. Structural organization of ABC transporters

In recent years, several ABC transporters have been crystallized in different conformations, providing a profound insight into their transport mechanism. However, further studies regarding the mechanisms of transport specificity are still required: how does the differentiation between similar ligands by the substrate binding protein is performed?

As mentioned before, the ABC transporters family hallmark is the core with the same modular architecture, where the NBDs are highly conserved in structure and sequence. In contrast, the

TMDs do not share significant sequence similarity and differ in the overall fold. Based on TMDs folds, so far, four types of ABC transporters have been identified: types I-III ABC importers (or energy coupling factor (ECF) transporter for type III) and exporter. The exporter fold architecture is present in both prokaryotes and eukaryotes, while the three ABC importers are exclusive for prokaryotes[104]. The first crystals structures of ABC transports were published in 2002 and 2006, for an intact type II ABC importer for vitamin B12 (BtuCD) from *E. coli*[114] and for a *Staphylococcus aureus* multidrug efflux pump exporter (Sav1866)[102], respectively. In 2007, Hollenstein *et al* published the first example of a type I importer, which is also the first structure of a complex between an ABC transporter from *Archaeoglobus fulgidus*, ModBC, and its associated substrate binding protein, ModA[115] – Figure 2.2.

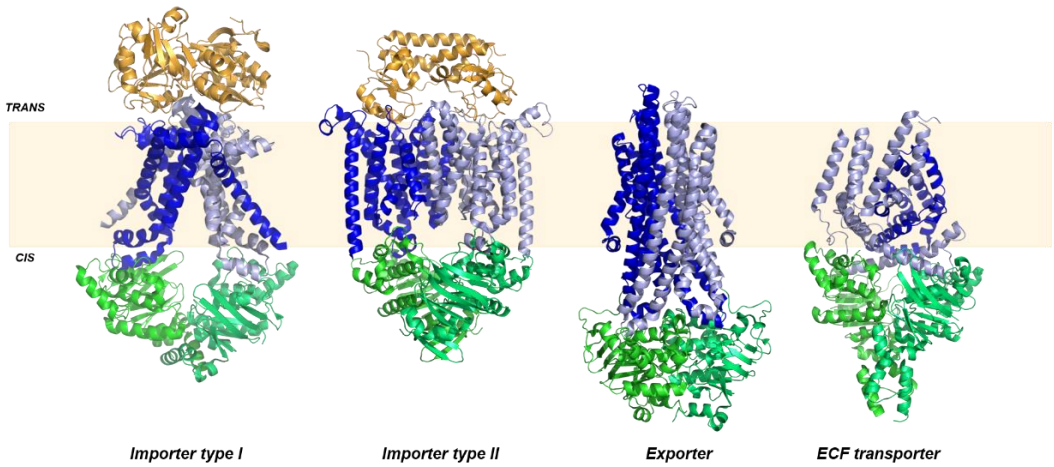


Figure 2.2. Cartoon representation of four distinct folds of ABC transporters. All share a general architecture with two transmembrane domains (blue and light blue) and two nucleotide-binding domains (green and lime green). In Type I and II importers, the ligands are captured by a substrate binding protein (orange) located in the periplasm (Gram-negative bacteria) or external space (Gram-positive bacteria and Archaea) and then delivered to the transmembrane domain. Importer type I: ModABC – PDB code 2onk[115]; importer type II: BtuCD+BtuF – PDB code 2qi9[116]; Exporter: Sav1866 – PDB code 2hyd[102] and ECF (energy coupling factor) transporter: folate ECF transporter – PDB code 4huq[117].

Independently of the ABC transporter type, the TMD switch between inward- (example: ModABC represented by the importer type I in Figure 2.2) and outward-facing (example: Sav1866 represented by the exporters in Figure 2.2) states, in an alternating access mechanism, which expose the ligand-binding site to the cis or trans side of the membrane[118], respectively. The *E. coli* importer BtuCD was the first example of the outward-facing conformation, revealing a 10 + 10 transmembrane helix topology in the TMD[114]. The same conformation was obtained for the crystal structure of the exporter Sav1866, which revealed a 6 + 6 transmembrane topology with a very different organization from BtuCD. Another different feature for the exporter is the long distance between the TM helices, placing the NBDs at ~20–30 Å away from the membrane[102]. Several structures that were solved among the years reinforce that ABC transporters exhibit strikingly different topologies in the TMDs, despite the conservation of the NBDs.

2.1.2.1. Substrate-binding proteins

Substrate-binding proteins (SBPs) are a class of proteins that are often associated with membrane protein complexes for transport or signal transduction[119]. Originally, they were only associated with prokaryotic ABC transporters[105,119] but, more recently, the SBPs have been implicated in the interaction with secondary transporters, namely tripartite ATP-independent periplasmic (TRAP)[120] and the tripartite tricarboxylate transporters (TTT)[121]. Also in prokaryotes, some two-component regulatory systems use SBDs for signal recognition. They also play a role in prokaryotic gene regulation through a ligand-binding moiety fused to a DNA-binding domain. Furthermore, in eukaryotes, a subset of G-protein coupled receptors (GPCRs) and ligand-gated ion channels (LGI) are activated through ligand binding to their respective SBPs[122]. A schematic overview of membrane protein complexes containing SBPs is represented in Figure 2.3.

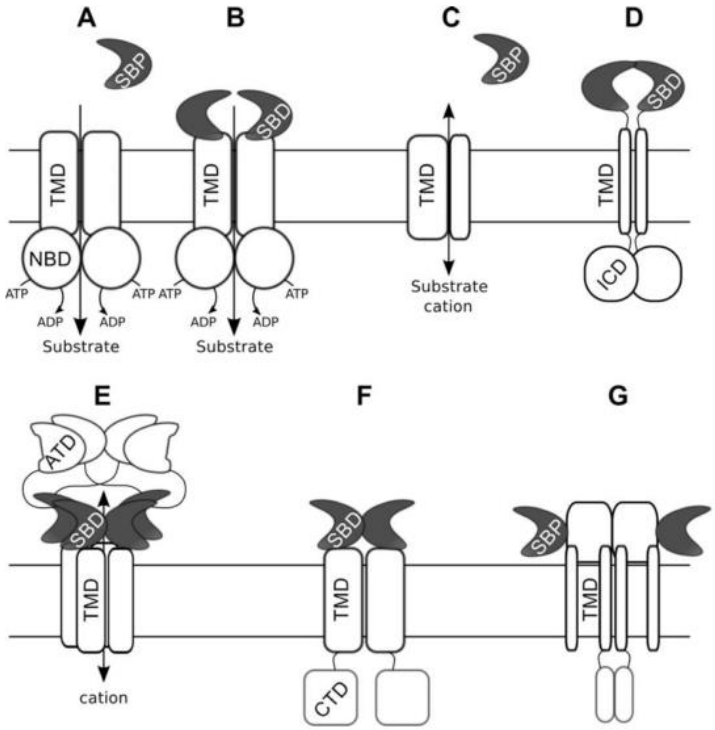


Figure 2.3. Schematic representation of SBP-dependent membrane proteins. A) ABC importer with the SBP in the periplasm or with a lipid-anchored SBP. B) ABC importer in a prokaryote with the SBD fused to the TMD, yielding two SBDs per transporter complex. C) TRAP transporter that can have either lipid-anchored or periplasmic SBP. D) Schematic of a Guanylate Cyclase-Atrial Natriuretic Peptide Receptor with an SBP, a single transmembrane helix and an intracellular domain (ICD). E) Ligand-gated ion channel, based on the ionotropic glutamate receptors (tetrameric structures), with at the top the ATD domains involved in the oligomerization of the protein. F) G-protein coupled receptor with a cytoplasmic domain (CTD). G) Schematic of a two-component sensor kinase. Adapted from [123].

The SBPs are responsible for the binding of different substrates with a wide range of binding affinities, that are often in the range of 0.01 to 1 µM. The high-affinity binding is essential for the efficiency of ABC transporters at low substrate concentrations. However, even at high substrate concentrations, the SBP is essential for the activation of the ABC transport system, as

demonstrated by Shuman[124] through the deletion of the gene encoding the maltose-binding protein (MBP). In this case, even when the other components involved in the transport system remain intact, the mutant is not able to grow on a high external concentration of maltose (25 mM) [124,125].

In Gram-negative bacteria, the SBP located in periplasm can be easily released by a cold osmotic shock, leading to the inactivation of transport due to the loss of the protein[126]. In prokaryotes without periplasm, Archae or Gram-positive bacteria, an N-terminal transmembrane segment is used to anchor the SBP to the cytoplasmic membrane or via an N-terminal lipid moiety. This is also true in some Gram-negative bacteria SBP, but a higher number of examples are needed for further demonstrating this hypothesis[127]. Regarding the ECF transporters, there are indications that they do not require an SBP, as described for the riboflavin transporter RibU from *S. aureus.* This protein was also the first crystal structure solved for this type of transporters, in 2010 by Zhang *et al*[128,129].

Due to the facility in isolating the SBP from the periplasm of Gram-negative bacteria, is not surprising that the first component structurally characterized of an ABC transport system was from this class: an L-arabinose binding protein from *E. coli* by Quiocho *et al* in 1974[130]. Currently, hundreds of structures are available at the PDB. Overall the SBPs are formed by two α/β domains, with a central β-sheet of five β-strands flanked by α-helices. Connecting the two domains (or lobes) is a hinge-region, with the ligand-binding site buried between the two lobes. In the absence of a ligand, the protein adopts an open conformation and upon ligand binding, by hinge bending motions, the closed conformation is stabilized and the ligand is trapped between the two domains (the "Venus fly trap" model[119]) – Figure 2.4. Possibly, after ligand binding the SBP is release to the TMDs is facilitated allosterically, with each domain interacting with one of the TMDs[104,131].

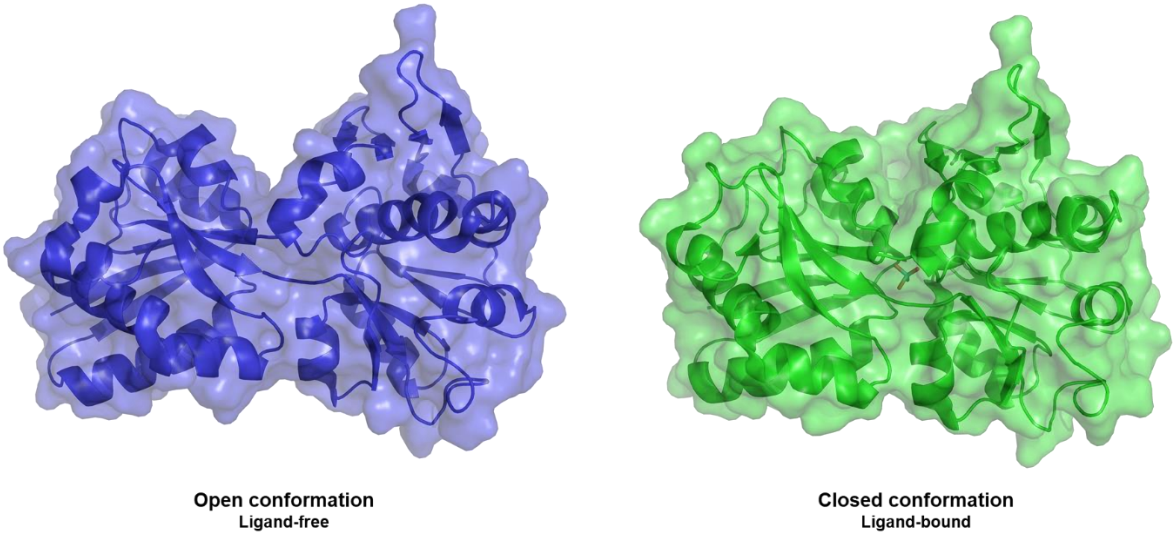


Figure 2.4. Representation of the rearrangements in ModA from *Methanosarcina acetivorans* upon ligand binding. In blue, the ligand-free form (PDB code 3k6v), with the binding site between two protein lobes accessible. In green, closed conformation upon the binding of molybdate (sticks) (PDB code 3k6w).

In 1999, Fukami-Kobayashi *et al* divided the SBP domain into three classes (or types) based on the topology of the β-sheets core, using crystallographic structures as reference[132]. In class I, the β-sheet topology of both lobes is $\beta_2\beta_1\beta_3\beta_4\beta_5$ with the hinge-region usually formed by three connecting strands. In class II, this region is formed by two connecting strands with a $\beta_2\beta_1\beta_3\beta_n\beta_4$ topology, where β_n represent the strand just after the first cross-over from the N-terminal domain to the C-terminal domain, and vice versa[123,132]. The MBP and molybdate-binding protein (ModA) from *E. coli* are examples of proteins from class I and class II, respectively. Also in 1999, with the determination of the crystallographic structure of a periplasmic zinc-binding protein TroA from *Treponema pallidum*, presented a single and long α-helical hinge region which lead to the formation of the class III of SBP in the Fukami-Kobayashi nomenclature system [133,134]. Eleven years later, in a deeper analysis, Berntsson *et al* [123] further categorized the SBPs in six clusters: from A to F– Table 2.1.

Table 2.1. Clusters of soluble SBPs based on Berntsson *et al* classification[123]. Adapted from[104].

Cluster	Type of ligands	Fukami-Kobayashi class[132]	Examples of proteins*	Ref
A	Metal ions (subcluster I) Siderophores (subcluster II)	III	*Escherichia coli* BtuF (2qi9) *Treponema pallidum* TroA (1toa)	116 133
B	Carbohydrates, Leucine, Isoleucine, Valine, autoinducer-2	I	*E. coli* MBP (1anf) *Vibrio harveyi* LuxP (1jx6)	135 136
C	Di- and oligopeptides, Arginine, cellubiose, nickel	II	*Lactococcus lactis* OppA (3drf) *Bacillus subtilis* AppA (1xoc)	137 138
D	Carbohydrates (subcluster I) Putrescine, thiamine (subcluster II) Tetrahedral oxyanions (subcluster III)	II	*Azotobacter vinelandii* ModA (1atg) *E. coli* TbpA (2qry)	139 140
E	Sialic acid, 2-keto acids, ectoine, pyroglutamic acid	II	*Halomonas elongate* TeaA (2vpo) *Bordetella pertussis* DctP6 (2pfz)	141 142
F	Trigonal planar anions, unknown ligands (subcluster I) Methionine (subcluster II) Compatible solutes (subcluster III) Amino acids (subcluster IV)	II	*E. coli* ProX (1r9l) *E. coli* HisJ (1hpb)	143 144

*Values in parenthesis correspond to PDB code.

Cluster A consists in SBPs classified as class III by Fukami-Kobayashi and all are involved in metal binding. The key feature of this group is the existence of a α-helix which acts as a hinge between both lobes. A small movement of both lobes upon substrate binding is observed [116]. Class I SBP forms the cluster B, with the N- and C-terminal located in different lobes and with three hinges between the two domains. The SBPs from this cluster interact with ABC-transporters, two-component histidine-sensory complexes, and guanylate cyclase-atrial natriuretic peptide receptors. The cluster C consists of class II SBPs that interact with ABC-transporters. They include an extra domain to accommodate their large ligands, such as di- and oligopeptides or cellobiose. Like the previous case, cluster D exclusively contains SBPs belonging to class II. The

distinct feature of this cluster is the short (4–5 amino acids) hinge-region formed by two strands and the variety of substrates. All substrate-binding proteins from cluster E are members of the tripartite ATP-independent periplasmic (TRAP) transporter family and classified as class II by the Fukami-Kobayashi system. These proteins use an electrochemical gradient to fuel the active translocation of substrates. The singular feature of this family is a large single β-strand that is part of each of the 2 five-stranded β-sheets of the lobes. Identical to cluster D, cluster F contain exclusively class II SBPs, with the two lobes connected by a significantly longer two segments hinge (8–10 amino acids), allowing a more flexibility between the open and closed conformation[123].

SBPs share a similar overall structure, but they have relatively low sequence similarity. For example, the SBP OpuAC orthologs in *Bacillus subtilis*[145] and *Lactococcus lactis*[146] (both Gram-positive bacteria) play the same role in the binding of glycine betaine and proline betaine. However, the sequence identity between the two proteins is about 43%. Otherwise, from the superposition of both structures results in an RMSD (Root-Mean-Square Deviation) of 1.09 Å for the 144 atoms aligned, demonstrating the conservation of the secondary elements. Considering the classifications described above – Table 2.1 – with several examples of SBPs with a similar structure that do not necessarily bind the same ligands, this might be a case of divergent evolution[123].

2.1.2.2. Transmembrane domain

In ABC transporters, the topology of the transmembrane domains (TMD) can be considered the 'fingerprint' of this type of systems since they generally do not display significant sequence conservation. Regarding the variability, the TMD share a similar topology within a transporter class with four distinct sets of folds currently recognized, wherein all the TMD constitute a translocation pathway. Depending on the fold topology, the TMD are designated type I ABC importer, type II ABC importer, ECF transporter (or type III ABC importer) and ABC exporter folds (Figure 2.2)[104]. The ABC exporters are not within the scope of this Thesis.

Type I ABC transporters are responsible for the uptake of ions, sugar, and other substrates, previously selected and captured by specific SBP[125]. The molybdate/tungstate transporter ModABC from *A. fulgidus* and the maltose transporter MalFGK from *E. coli* are examples of this type of importers. Here, the two TMD can be homodimers or heterodimers, with a core membrane topology of five transmembrane helices per TMD. In some cases, an additional N-terminal transmembrane helix wraps around the partner TMD [101,104,147]. The transport cycle (Figure 2.5a) starts with the substrate capture by SBP. When the substrate-SBP complex occurs, the ABC transporter is in an inward-facing conformation – resting state. During the transient and low-affinity interaction between SBP and TMD, the NBD adopts a closed conformation after ATP binding, which converts the TMDs into an outward-facing conformation. The substrate is then transferred to an occluded cavity formed between the SBP and the translocation pathway located halfway

across the membrane. ATP binding is crucial to stabilizing the interaction between SBP and TMD and allows the SBP to deliver the ligand to the TMD component. Once the substrate binds the TMD, the ATP can be hydrolyzed, the NBD opens with the concomitant dissociation of SBP. The TMD re-orient to an inward-facing conformation – returns to the rest state, releasing the ligand from the low-affinity site to the cytoplasm. This structural transition between two states is known as the 'alternate access' model described by Jardetzky in 1966[104,108,148].

The bacterial type II ABC importers are involved in uptake of metal chelates, such as cobalamin, heme and other iron-containing complexes[149]. The mechanism proposed for the *E. coli* BtuCD-F transporter for vitamin B12 serves as a model[114]. These importers contain 10 aligned transmembrane helices per each identical TMD. As described for type I, these transporters are also depending on SBP to select and capture the ligands. In the absence of a loaded SBP, the TMD is an outward-facing conformation. Independently of ATP binding, a high affinity and highly stable complex is formed between the loaded SBP and the ABC transporter. After ligand transfer between SBP to the translocation pathway between the TMDs, the ATP is hydrolyzed and the ADP and phosphate are released, which requires the (at least partial) opening of the NBD dimer. Because of the size of the trapped substrate, the TMD is unable to close, which may cause a strained conformation and induce pressure into the substrate. This may contribute to substrate dissociation into the cytoplasm. Upon substrate release, ABC transporter collapses into an asymmetric conformation that no longer features a central cavity, due to the TMD close conformation [116,149]. The transport cycle for type II remains somewhat unclear but the overall steps are illustrated in Figure 2.5b.

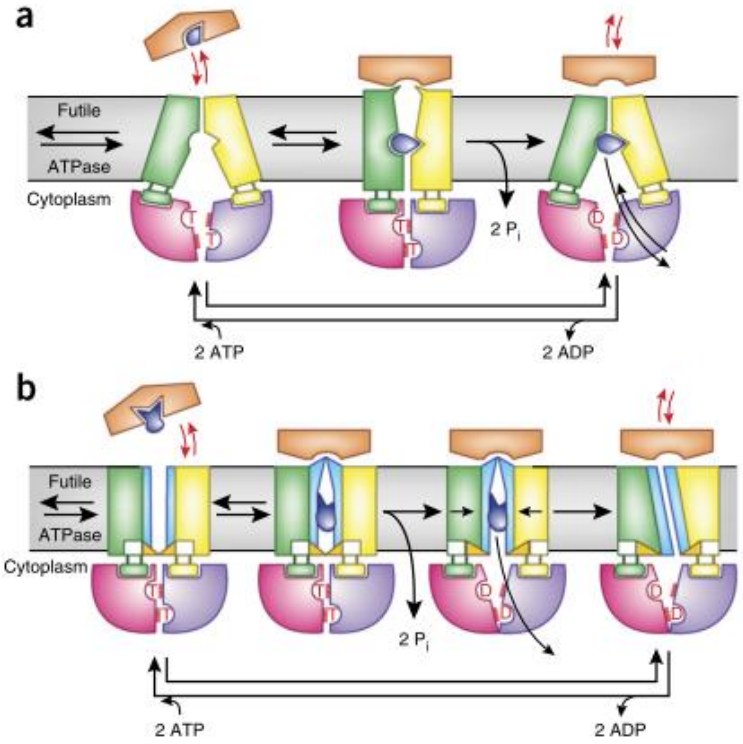


Figure 2.5. Schematic representation of the mechanisms of type I (a) and type II ABC importers (b). Adapted from [149].

Exclusively found in prokaryotes, a unique family of ABC transporters is responsible for the uptake of micronutrients (such as vitamins, metals, and amino acids) – the ECF transporters. These transporters consist of two functional and structural unrelated TMD: a transmembrane coupling subunit, called T-component (or EcfT subunit) and a membrane-embedded substrate-binding subunit, designated the S-component (EcfS). The T-component core is formed by four to eight TM helices and the S-component by six TM helices, with few S-components presenting an additional N-terminal helix[150]. The S-component interacts extensively with the EcfT and poorly with the NBD and it is responsible for a high-affinity recognition of substrate, discarding the necessity of an SBP. The T-component links the EcfS to the NBD, which is form by EcfA and EcfA' subunits [104,128,149].

Recently, Karpowich *et al* proposed a transport cycle based on the ECF transporter for riboflavin from the pathogen *Listeria monocytogenes*, LmECF–RibU[151]. It combines the binding of ATP to EcfAA' dimer that adopts a close conformation and sliding the EcfT coupling helices apart. This conformational change disrupts the interaction with EcfS, that assumes an extracellular-facing orientation in the membrane. Upon ligand binding to EcfS, this complex interacts with the EcfTAA' module and, after the ATP hydrolysis, the substrate is released into the cytoplasm. As such, in contrast with the other importers, the ECF transporters does not open a translocation pathway - Figure 2.6. Despite the proposed mechanism, more structures of ECF transporters need to be solved to clarify its mechanism and transport cycle.

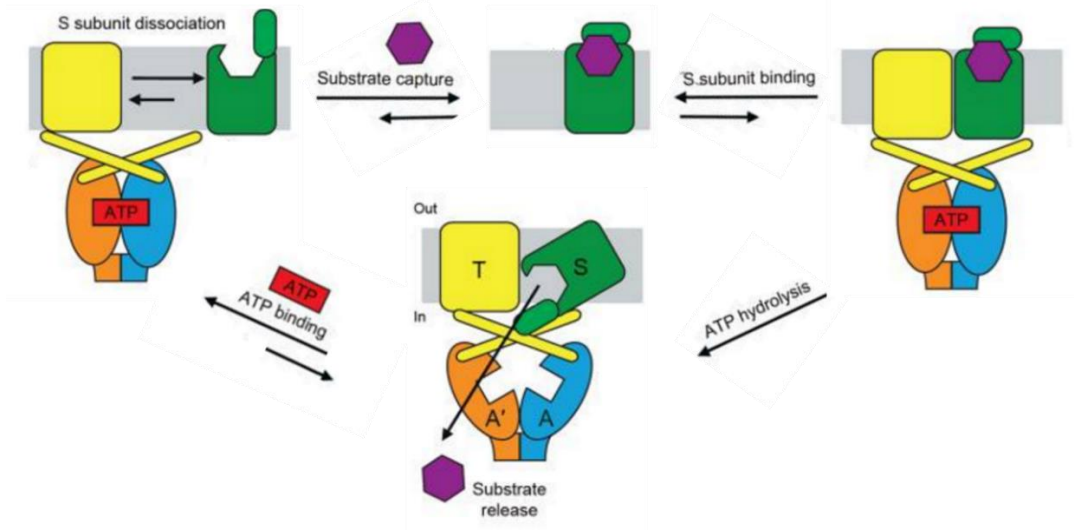


Figure 2.6. Schematic representation of the mechanisms of energy-coupling factor (ECF) transporters. Adapted from[151].

2.1.2.3. Nucleotide binding domain

The nucleotide-binding domain (NBD) acts as the engine of ABC transporters associated with the ATP-binding and hydrolysis. In ABC transporters, the NBD are a subgroup of the diverse superfamily of P-loop NTPases that depend on magnesium ions for catalysis[152]. Each NBD is composed by two subdomains: the RecA-like domain, also found in other P-loop ATPases, and a α-helical domain, which is unique to ABC transporters. Despite the described differences in SBP and TMD, the NBD are highly conserved and contain several motifs, including: 1) Walker A (P-loop), a phosphate binding loop with a highly-conserved lysine (GXXGXGK(S/T) motif); 2) Walker B (ΦΦΦΦDE, where Φ is a hydrophobic residue) involved in the Mg^{2+} coordination; 3) A-loop, usually contains a tyrosine (or other aromatic residue) that helps to position the ATP; 4) D-loop (SALD motif), involved in the formation of the ATP hydrolysis site; 5) H-loop, contains a highly conserved histidine that helps the positioning of the attacking water molecule and the Mg^{2+} ion; 6) Q-loop (named due to a conserved glutamine residue), is located at the interface between the RecA-like domain and the α-helical domain and is a major site of interaction with the TMDs; and finally 7) The ABC-binding signature motif (LSGGQ), located in the α-helical domain and used to identify NBD containing proteins in protein databases. The NDB dimers have two ATP binding/hydrolysis sites between the Walker A motif of one monomer and the LSGGQ motif of the other, switching between a closed and open conformation. Results obtained by Patzlaff *et al*[153] supports a transport cycle stoichiometry of 2:1 (ATP molecule to transported substrate), although it is not possible to affirm that this stoichiometry is conserved among all ABC transporters [104,149,154]. Upon ATP hydrolysis, Pi and ADP are released. This leads to the destabilization of the dimer allows the NBDs to move apart and, concomitantly, the delivering of the substrate into the cytoplasm.

As mentioned previously, ABC transporters are crucial throughout the evolution of life, from the most complex organism to a single cell. The following section discusses the role of these machines in the transport of molybdate and tungstate in prokaryotes.

2.1.3. Bacterial transporters for tungstate and molybdate

2.1.3.1. General description

Molybdenum and tungsten are essential metals in all types of life, being imported by cells in the form of their soluble oxyanions (MoO_4^{2-} or WO_4^{2-}) by highly specific ABC transporters[155,156].

In bacteria, there are three classes of tungstate/molybdate ABC transporters: ModABC[157], TupABC[18] and WtpABC[158]. These are examples of the type I ABC transporters. They are composed of a periplasmic substrate-binding protein (component A), two transmembrane domains (component B) and two nucleotide-binding protein/ATPases (component C)[156]. The

genes encoding the three components are organized in an operon (*mod*/*wtpABC*) and regulated by a transcription factor known as ModE (in the case of the ModABC operon)[159,160]. The ModABC-ModE systems are widespread in prokaryotes, but not ubiquitous. Variations of ModE-like proteins were also observed in other Moco-dependent organisms. On the other hand, regulation mechanism of WtpABC and TupABC transporters is still unclear[13,159].

In biological systems, the incorporation of tungsten, as a cofactor, is limited to bacteria and archaea, with the W-containing enzymes almost restricted to obligate anaerobic prokaryotes[17,161]. In contrast, the similar element molybdenum can be also found in several enzymes from eukarya, such as in the active site of the human aldehyde oxidase[162]. In most cases, the function of Mo and W-enzymes is to catalyze oxygen atom transfer reactions.

Due to the similarities between the two elements (for more details see section 1.1.), one of the challenges of biological systems is to differentiate one from the other and avoid the incorrect insertion in the catalytic site of enzymes. Misincorporation of metals often leads to compromised activity or even inactive enzymes[163–165]. Probably, the mechanism to avoid this misincorporation is dependent on the *modA*/*tupA*/*wtpA* genes that encode a periplasmic molybdate/tungstate-binding protein (protein A), that acts as a first selection gate. The ModA, WtpA, and TupA proteins differ in their binding affinity for molybdate versus tungstate, primary sequence and oxyanion coordination chemistry[156,158,166–168]. The dissociation constants reported to date suggest that TupA and WtpA are strongly selective for tungstate whereas ModA cannot discriminate between molybdate and tungstate[156].

Several crystal structures of ModA/WtpA proteins are available in PDB, namely Mod/WtpA from *Pyrococcus furiosus*[169], ModA from *E. coli*[170] and ModA *A. fulgidus*[115]. In contrast, TupA has been poorly studied. Despite the existence of the crystal structure of *Geobacter sulfurreducens* deposited in the PDB (PDB code 3lr1) with 46% of identity, the TupA transporter system was only characterized biochemically in *Eubacterium acidaminophilum*[166,171] and *Campylobacter jejuni*[172]. Atomic detail of tungstate-binding proteins is crucial to understand ligand coordination and the mechanism of metal differentiation, especially if biotechnological applications are considered.

In eukaryotic organisms, the studies regarding the transport of molybdate are quite limited, with only two high-affinity molybdate transporters identified: Molybdate Transporter type 1 (MoT1) and type 2 (MoT2)[173,174]. MoT1 has been identified in higher plants, algae, fungi and bacteria, and was firstly described for *Arabidopsis thaliana* and *Chlamydomonas reinhardtii*[173]. It belongs to the sulfate transporter superfamily and is involved in molybdate transport/accumulation. MoT2 was firstly identified in *C. reinhardtii* but is also present in other algae and animals (including mammals). More studies need to be performed to clary the mechanism and cellular localization of these two transporters [7].

In prokaryotes and eukaryotes, once the oxyanion reaches the cytoplasm is incorporated in the biosynthesis the molybdenum (Moco) or tungsten (Woco) cofactors, which play a crucial role in the activation of several enzymes. This topic is described in more detail in section 3.1. of chapter 3.

2.1.3.2. Why study tungstate/molybdate ABC transporters in *Desulfovibrio alaskensis* G20?

Tungsten and molybdate are essential elements for the Gram-negative sulfate-reducing bacteria (SRB), *Desulfovibrio alaskensis* G20 (*Da*G20, formerly *D. desulfuricans* G20). The mesophilic G20 strain was isolated from an oil well in Ventura County, California and derived from *D. desulfuricans* G100A. *Da*G20 is a nalidixic acid resistant derivative that obtains energy from sulfate reduction and produces hydrogen sulfide, which is chemically reactive and toxic to plants, animals, and humans[175,176].

In anoxic conditions (e.g., in oil and gas pipelines), SRBs are the main cause of microbially influenced corrosion (MIC), either through electron withdrawal reactions or production of corrosive chemical species as hydrogen sulfide. MIC is a degradation process with severe economic consequences in different industries including shipping, oil and energy, aviation, chemical, paper, and wastewater infrastructures. Although diverse strategies have been studied to control the biogenic production of sulfide, the main constraints are the environmental regulations. In the shipping industry, the demand for better anti-fouling paints is motivating the scientific community to develop new agents using bio and nanotechnology. The biggest challenge in the field is to create easily degradable biocides, which are non-toxic to non-target organism[177].

Several studies demonstrated that molybdate and nitrite can act as metabolic inhibitors of SRB. Nitrite blocks the sulfate respiration by inhibiting the dissimilatory sulfite reductase, which catalyzes the reduction of sulfite to sulfide in all SRB. Molybdate, as sulfate analog, can be converted to adenosine phosphomolybdate by the ATP sulfurylase. This product is unstable and hydrolyzes spontaneously to AMP and molybdate, depleting cellular ATP reserves. Due to the broader activities of the biocides (e.g., glutaraldehyde, benzalkonium chloride or bronopol) used in these situations, the synergy between the two approaches may lead to a decrease of the microbial resistance, decreasing the costs and environmental toxicity[178,179].

An analysis of the *Desulfovibrio* genome annotated to date shows that ABC transporters for molybdate and tungstate are encoded in the chromosomes of these organisms and belong to the Mod and Tup family of proteins, respectively[180]. In 2015, Nair *et al*, observed that *Da*G20 cultures exposed to high concentration of molybdate affect the expression of proteins involved in energy metabolism, ion transport, cofactor insertion and other cellular mechanisms, and downregulate the two SBPs: ModA, and TupA[181].

In this project, our goal is to characterize the mode of action of tungstate/molybdate ABC transporters (TupABC and ModABC) using different techniques such as X-ray crystallography, isothermal titration calorimetry and SAXS. Especially, we aim to structurally understand how the SBP distinguishes two similar substrates to support further studies of anti-MIC targets and, even, the development of tungstate biosensors.

2.2. Experimental procedure

The work presented in this section results from a collaboration with Dr. Maria Gabriela Rivas from Universidad Nacional del Litoral, Argentina. The DNA cloning and site-directed mutagenesis protocol where the *tupA* gene from *Desulfovibrio alaskensis* G20 (*Da*G20) was used as a template to produce three mutants (*tupA*_R118K, *tupA*_R118Q and *tupA*_R118E), was performed by Raquel S. Cordeiro and Dr. Rashmi Nair from UCIBIO@REQUIMTE, FCT-UNL.

The DNA cloning, protein expression and purification of the protein ModA from *Da*G20 (UniProt Q30VI5) was carried out by Dr. Rashmi Nair from UCIBIO@REQUIMTE, FCT-UNL.

The reagents were purchased from Sigma-Aldrich unless stated otherwise.

2.2.1. Protein expression and purification

2.2.1.1. Tungstate binding protein - TupA

The *tupA* gene (UniProt Q316W1) was amplified from the *Da*G20 genome and cloned into the pET-46 Ek/LIC vector using the Ek-LIC cloning system (Novagen), as presented in [13]. To confirm the successful cloning of the TupA with a histidine tag (MA**HHHHHH**VDDDDKMLEVLFQGP) attached to the N-terminal, Sanger sequencing method (STAB VIDA, Portugal) was performed using T7 forward and reverse primers.

The protein was heterologously expressed in *Escherichia coli* BL21(DE3) cells. Several expression conditions were tested by varying the induction temperature (293 or 310 K), induction time (3, 5 or 16 hours) and isopropyl β-D-thiogalactopyranoside (IPTG, from NZYtech) concentration (0, 0.2, 0.5 and1.0 mM). The highest expression yield was obtained using 1 mM IPTG at 293 K during 3 hours of induction, in Luria-Bertani (LB) medium. The cells were collected by centrifugation at 8656 × g for 15 min, washed in 5 mM Tris-HCl pH 7.6, centrifuged and resuspended again in the same buffer containing DNaseI (5 µg/mL) at a concentration of 2 g cells/mL. The cells were disrupted by sonication (UP100H, Hielscher Ultrasonics) and the crude extract centrifuge at 15557 × g for 1 hour (Eppendorf, Centrifuge 5804 K).

The chromatographic purification protocol for TupA is comprised of two main steps [13]. First, the soluble fraction was loaded into a DEAE Sepharose Fast Flow (GE Healthcare Bio-Sciences AB) column equilibrated with 3 column volumes (CV) of 5 mM Tris-HCl pH 7.6. TupA was eluted using a gradient from 5 mM to 500 mM Tris-HCl pH 7.6 in 8 CV, and all the fractions containing the protein were pooled and concentrated by ultrafiltration using an Amicon Pro Purification System (EMD Millipore). The sample was afterward injected into a Superdex 75 HR10/300 GL column (GE Healthcare Bio-Sciences AB) and equilibrated with 50 mM potassium phosphate pH 7.6 and 150 mM NaCl. The fraction containing the pure protein was dialysis overnight against 5 mM Tris-HCl pH 7.6, concentrated up 7.5 mg/mL (determined by ultraviolet absorption at 280 nm, using

the ε_{280nm} of (29700 ± 700 $mM^{-1}.cm^{-1}$ [13]) and stored at -80 °C until further use. The purification yield was calculated to be approximately 10 mg of soluble protein per liter of cell culture.

2.2.1.2. TupA mutants of the arginine 118

Based on sequence analysis, the R118 was pointed as an important amino acid for oxyanion coordination. To understand the role of the said amino acids, site-directed mutagenesis was performed resulting in the mutants: TupA_R118K, TupA_R118Q and TupA_R118E. The first approach to express the proteins was by testing the TupA wild-type expression conditions. To overcome the high percentage of insoluble protein obtained in section 2.2.1.1, different strains of *E. coli* were tested as expression hosts (Rosetta 2(DE3)pLysS, Origami (DE3) and Tuner (DE3), from Merck Millipore), different induction temperatures (292 K or 303 K) and IPTG concentrations (0.1, 0.3, 0.5, 0.8 and 1 mM). The highest expression yield condition for each mutant is summarized in Table 2.2. The three mutants were grown in sterile LB medium supplemented with ampicillin (100 µg/mL) at 310 K and 180 rpm. Protein expression was induced at OD_{600nm} = 0.6 for 16 hours at 292 K.

Table 2.2. Growth conditions with the highest expression yield for TupA mutants.

TupA mutant	Expression strain	Induction temperature	IPTG concentration (mM)
R118K	*E. coli* Rosetta 2		1.0
R118E	*E. coli* Origami	292 K	1.0
R118Q	E. coli Tuner		0.3

The cells were harvested (5000 × g for 20 min) and resuspended in 50 mM phosphate buffer pH 8.0 with 500 mM NaCl, DNaseI (5 µg/mL) and 1 tablet/L of Protease Inhibitor Cocktail – EDTA free. The cells were disrupted by sonication and centrifuged at 15557 × g for 1 hour at 277 K.

In this case, the purification was performed in one-step using an immobilized metal affinity chromatography (IMAC), His GraviTrap (GE Healthcare) column, following the manufacturer's instructions. The proteins were eluted using 50 mM sodium phosphate buffer pH 8.0 containing 500 mM NaCl and 250 mM imidazole. Fractions containing pure TupA_R118K, TupA_R118E and TupA_R118Q were dialyzed against 5 mM Tris-HCl pH 7.6 during 16 hours at 277 K.

2.2.2. Protein crystallization and X-ray diffraction experiments

2.2.2.1. TupA crystals, data collection and processing

The first crystallization trials for *Da*G20 TupA were performed using protein at 7.5 mg/mL in 5 mM Tris-HCl pH 7.6 and the Oryx8 protein crystallization robot for nanocrystallization setups (Douglas Instruments). More than 384 different crystallization conditions were tested, using several commercial screens, namely the PEG/Ion HT (Hampton Research), the JBScreen Classic 1-10 (Jena Bioscience) and in-house sparse matrix screen (based on the screen of Jancarik *et al*[182]– formulation in the Table A1, Appendix). After 4 days at 293 K, the first crystallization hit was obtained using the sitting-drop vapor diffusion method (SWISSCI 'MRC' 2-Drop Crystallization Plates – 96 wells, Douglas Instruments), in a 1 µL drop (containing 50% protein). The crystallization condition was composed of 30% (w/v) polyethylene glycol (PEG) 3350 as precipitant, 0.2 M magnesium chloride and 0.1 M HEPES pH 7.5 as a buffer, from the in-house screen. To improve the quality and increase the size of the colorless crystals, scale-up and optimization experiments were performed, using the hanging-drop and sitting-drop vapor diffusion methods. The optimization attempts were performed by varying the precipitant concentrations (from 26 to 34% (w/v) PEG 3350), the drop proportion and temperature (293 K or 277 K). A single crystal, with maximum dimensions of 0.3 × 0.15 × 0.06 mm, was obtain after 3 days (at 293 K) in a 2 µL of protein (at 7.5 mg/mL) + 1 µL of the precipitant solution (30% (w/v) PEG 3350) drop (Figure 2.7).

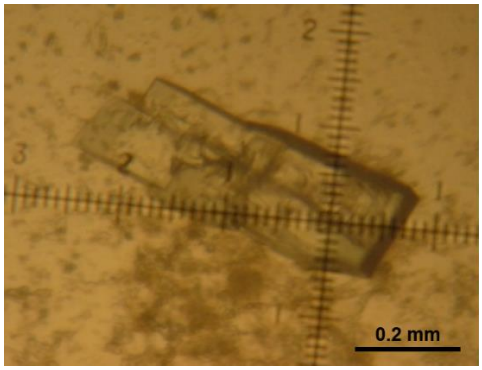


Figure 2.7. Crystal of TupA protein from *Desulfovibrio alaskensis* G20. The crystal was obtained with 30% (w/v) polyethylene glycol 3350, 0.2 M magnesium chloride and 0.1 M HEPES pH 7.5.

The crystal was transfer into a new drop with a harvesting solution containing a higher percentage of PEG 3350 (32% (w/v). After 3-5 min incubation, the crystal was then flash-cooled in liquid nitrogen, using paratone oil as a cryoprotectant, before transfer to a gaseous nitrogen stream for data collection. A complete dataset was collected at beamline ID23-1 from the European Synchrotron Radiation Facility (ESRF, France), at a wavelength of 0.954 Å on a PILATUS 6M-F detector. The crystal diffracted beyond 1.40 Å resolution and belonged to space group $P12_11$,

with unit cell dimensions a = 52.25 Å, b = 42.50 Å, c = 54.71 and β = 95.43 º. The calculated Matthews coefficient[48] was 2.09 Å^3/Da, suggesting the presence of one monomer per asymmetric unit with a solvent content of 40.84%.

The collected dataset was integrated with the program XDS package[183] and scaled with AIMLESS[184] from the CCP4 suite of programs [185]. The data collection and processing statistics are presented in Table 2.3.

Table 2.3. X-ray crystallography data-collection statistics for TupA crystal. Values in parenthesis correspond to the highest resolution shell.

Data collection parameters	
X-ray Source	ID23-1 (ESRF, France)
Detector	PILATUS 6M-F
Wavelength (Å)	0.954
Processing statistics	
Unit-cell parameters (Å, ˚)	$a = 52.32$ $b = 42.53$ $c = 54.75$ $\beta = 95.45$
Space group	$P12_11$
Molecules per AU	1
Matthews coefficient (Å^3/Da)	2.09
Mosaicity (˚)	0.22
Resolution range (Å)	42.53-1.40 (1.42-1.40)
$< I/\sigma I >$	10.3 (2.1)
R_{pim} **(%)**[a]	3.3 (28.8)
Multiplicity	2.9 (2.9)
Number of observed reflections	135886 (6650)
Number of unique reflections	46519 (2277)
$CC_{1/2}$	0.998 (0.903)
Completeness (%)	99.3 (98.5)

[a] $R_{pim} = \sum_{hkl}[\frac{1}{N-1}]^{1/2} \sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hlk}\sum_i I_i(hkl)$, where N is the multiplicity measured.

To obtain a crystal structure of the holo form of TupA, with tungstate or molybdate bound, soaking and co-crystallization experiments were performed. For the soaking experiments, crystals obtained in the described condition were stabilized by adding a harvesting buffer solution containing 32% (w/v) PEG 3350. The crystals were then incubated with a 5, 10 or 20-fold excess of ligand (prepared in 0.2 M magnesium chloride, 0.1 M HEPES pH 7.5 and 32% (w/v) PEG 3350) for 10 min up to 24 hours. Macroscopically, no damage was observed on the crystals during the

soaking and were flash frozen using the mentioned cryo-protectant. More than 120 crystals were tested, although the majority had poor to non-existing diffraction.

Additionally, the co-crystallization methodology was followed. The protein (15 mg/mL) was incubated with the mentioned ligands in a 10-fold excess for 25 to 60 min and then passed through a PD MiniTrap G-25 column (GE Healthcare), in order to remove the unbound ligand prior to the crystallization trials. Several crystallization screens were tested for TupA (at 7.5 mg/mL) at 293 K, using the nanocrystallization setup on Oryx 8 crystallization robot and the sitting-drop vapor diffusion method (96-well plates). No crystallization hits were obtained despite more than 480 different conditions tested. Due to the high percentage of clear drops that resulted from this approach, microseeding was tried using the apo-TupA crystals, but no good diffracting crystals were obtained.

2.2.2.2. Crystallization of TupA mutants and data collection

Initial crystallization trials of TupA mutants (TupR118K - 14 mg/mL, TupR118E - 10 mg/mL and TupR188Q - 11 mg/mL) were performed using nanocrystallization setup in Oryx8 protein crystallization robot. Drops of 0.5 µL of protein (in 5 mM Tris-HCl pH 7.6 buffer) + 0.5 µL of the reservoir solution were set up using the sitting drop vapor diffusion method, at 277 K. Several crystallization screenings were used: an in-house sparse matrix screen (Appendix, Table A1), Wizard Classic 1 and 2 (Molecular Dimensions/Rigaku), MemStart (Molecular Dimensions), among others. The first crystallization hit was obtained after 3 days, in the same condition for the three mutants, where colorless plate-shape crystals appeared: 0.1 M magnesium chloride, 0.1 M MES (pH 6.5) and 30% (w/v) polyethylene glycol 8000, from the in-house sparse matrix screen. Crystals were flash-cooled directly in liquid nitrogen prior to transfer to a gaseous nitrogen stream using paratone oil as a cryoprotectant. The crystals were tested at I24 beamline at DLS synchrotron but none diffracted. Several attempts have been made to overcome this problem but, so far, unsuccessfully.

2.2.2.3. ModA crystals, data collection and processing

The first crystallization trials were performed using *Da*G20 ModA at 10 mg/mL in 5 mM Tris-HCl pH 7.6. The in-house sparse matrix screen (Appendix, Table A1) was tested using the hanging-drop vapor diffusion method in a 24-well XRL plate (Molecular Dimension), with 1 µL of protein + 1 µL of the precipitant solution and 700 µL of precipitant solution in the reservoir. After two days at 293 K, several needle-like crystals grew in different conditions. The most promising crystals were found in the following conditions: **A)** 1 M potassium/sodium tartrate and 0.1 M MES pH 6.5; **B)** 0.2 M lithium sulfate, 0.1 M citrate buffer pH 5.5 and 30% (w/v) PEG 3350; **C)** 0.1 M ammonium sulfate, 0.1 M citrate buffer pH 5.5 and 30% (w/v) PEG 8000; **D)** 0.2 M ammonium sulfate, 0.1 M cacodylate buffer pH 6.5 and 30% (w/v) PEG 6000; **E)** 0.2 M ammonium acetate, 0.1 M citrate

buffer pH 5.5 and 30% (w/v) PEG 3350; and **F)** 0.2 M ammonium sulfate and 30% (w/v) PEG 3350. These conditions were optimized by varying the precipitant percentage, the drop proportion and crystallization temperature. The best crystals derived from the optimization procedure were harvested and flash frozen using paratone as a cryo-protectant. Several cryo-protectant solutions containing glycerol (15 and 25%) and PEG 400 concentration (15%) were also tested. The crystals were tested in beamline BM30 at the European Synchrotron Radiation Facility – ESRF (France), diffracting between 3.5 and 5.0 Å resolution. Later, for the crystals grown in the presence of 28% (w/v) PEG 6000 (optimization of condition D – Figure 2.8), a complete dataset was collected with a 3.0 Å resolution and processed in $P3_121$ space group, using XDS and AIMLESS (Table 2.4 and Figure 2.9).

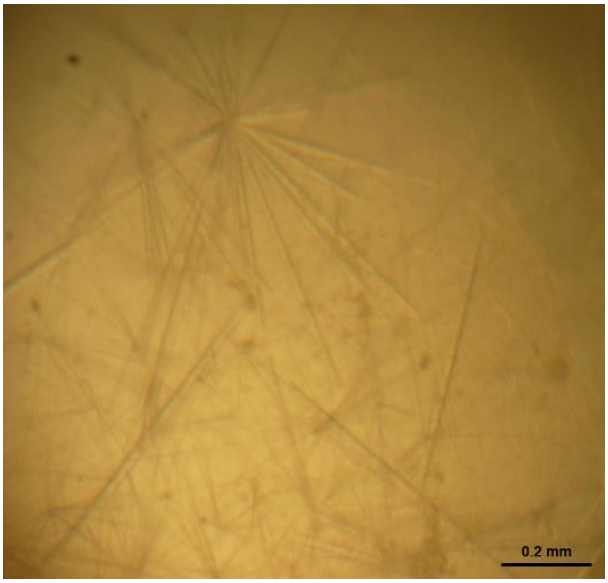


Figure 2.8. Crystals of ModA protein from *Desulfovibrio alaskensis* G20. The crystals were obtained with 0.2 M ammonium sulfate, 0.1 M cacodylate buffer pH 6.5 and 28% (w/v) PEG 6000.

New attempts were carried out to improve the crystal quality and resolution, namely by adding crystallization additives. In a 24-well plate, drops containing 1 µL of protein (10 mg/mL) + 0.8 µL precipitant solution (0.2 M ammonium sulfate, 0.1 M cacodylate buffer pH 6.5 and 28% (w/v) PEG 6000) + 0.2 µL of additive (Additive Screen, Hampton Research), and 700 µL of precipitant solution in the reservoir were setup at 293 K. After 6 days, colorless crystals appeared in the presence of 2% (w/v) of benzamidine hydrochloride. The best crystals were harvested in a solution containing a higher precipitant percentage (30% (w/v) PEG 6000) and allowed to stabilize for 3-4 min before flash-cooled directly in liquid nitrogen with 15% glycerol as cryoprotectant. One dataset was collected in beamline I24 at the Diamond Light Source – DLS (Oxfordshire, United Kingdom). One of the tested crystals diffracted up to 2.77 Å resolution and the data was processed in $P12_11$ space group. The resolution improved in this case, but due to the difficulties in processing the dataset collected at DLS, the ModA structure determination was carried out using the dataset from BM30 beamline.

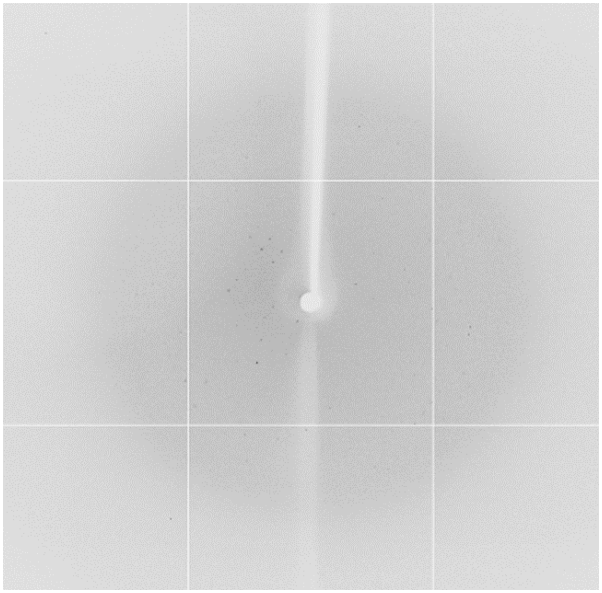


Figure 2.9. Diffraction pattern obtained at beamline BM30A (ESRF, France) for a ModA crystal. Resolution at the edge is 3.0 Å.

Table 2.4. X-ray crystallography data-collection statistics for ModA crystal. Values in parenthesis correspond to the highest resolution shell.

Data collection parameters	
X-ray Source	BM30A (ESRF, France)
Detector	ADSC Q315r CCD
Wavelength (Å)	0.9797
Processing statistics	
Unit-cell parameters (Å, ˚)	$a = 103.35$ $b = 103.35$ $c = 62.07$ $\gamma = 120.0$
Space group	$P3_221$
Molecules per AU	1
Matthews coefficient (Å³/Da)	3.54
Mosaicity (˚)	0.19
Resolution range (Å)	44.75-3.0 (3.18-3.0)
$< I/\sigma I >$	13.0 (2.7)
R_{pim} **(%)**[a]	6.9 (32.0)
Multiplicity	10.8 (11.1)
Number of observed reflections	85687 (14017)
Number of unique reflections	7917 (1260)
$CC_{1/2}$	99.7 (86.8)
Completeness (%)	100 (100)

[a] $R_{pim} = \sum_{hkl}[\frac{1}{N-1}]^{1/2} \sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hlk} \sum_i I_i(hkl)$, where N is the multiplicity measured.

2.2.3. Structure solution, model building and refinement of TupA

Structure determination was carried out by molecular replacement (MR) using PHASER from the CCP4 suite of programs[186]. Several molecular models were selected according to sequence alignment homologies, namely: a conserved functionally unknown protein from *Vibrio parahaemolyticus* serotype O3:K6 (PBD code 3muq) and the *Geobacter sulfurreducens* TupA (PDB code 3lr1), after omitting all the cofactors and solvent molecules. The MR solution could only be obtained when the two models were superimposed and small domains of the protein were used separately: Domain I, including the first 81 residues; Domain II complying residues 82 to 188; and finally, Domain III with residues 189 to 236.

After structure solution, Buccaneer software was used for the automated model building (Cowtan, 2006) and REFMAC 5 for restrained refinement [187]. The water molecules were automatically added by REFMAC 5 and manually inspected in COOT [188]. Geometrical validation and model improvement was carried out using PDB_REDO[189] and the validation tools of COOT. The final values of 17.6 and 21.7 for $\boldsymbol{R_{work}}$ and $\boldsymbol{R_{free}}$ factors were obtained, respectively. The Ramachandran plot has 97.15% of the residues in the preferred regions, without outliers. Mean bond angle and bond length deviations from ideal values and other refinement statistics are presented in Table 2.5. The deposited model contains 250 protein residues, 332 water molecules, two chlorides and one sodium ion.

Table 2.5. Structure refinement statistics for TupA.

Refinement statistics	
Resolution range (Å)	42.53–1.40
$\boldsymbol{R_{factor}}$ **(%)**[a]	17.6
$\boldsymbol{R_{free}}$ **(%)**[b]	21.7
Number of water molecules	332
Other heteroatoms	1 sodium, and 2 chlorides
Average B factor for all atoms (Å^2)	23.67
RMSD from ideal geometry	
Bond lengths (Å)	0.022
Bond angles (°)	2.017

[a] $R_{factor} = \sum ||F_{calc} - F_{obs}|| / \sum |F_{obs}| \times 100$, where F_{calc} and F_{obs} are the calculated and observed structure factor amplitudes, respectively.
[b] R_{free} is calculated for a randomly chosen 5% of the reflections.

Coordinates and observed structure factor amplitudes of *Da*G20 TupA have been deposited in the PDB under the accession code **5my5**.

2.2.4. Structure solution, model building and refinement of ModA

Structure determination was carried out by MR using PHASER from the CCP4 suite of programs[186]. According to sequence alignment homology, the ModA from *Azotobacter vinelandii*[139] (PBD code 1atg) was used as a template, after omitting all the cofactors and solvent molecules. As in the previous case, the MR solution could only be obtained when the template was decomposed into domains: Domain I, including the first 80 residues and between 192 to 232; and Domain II with residues 81 to 191.

After structure solution, Autobuild from PHENIX[190] was used for the automated model building and REFMAC 5 for restrained refinement[187]. The water molecules were automatically added by REFMAC 5 and manually inspected in COOT[188].

The structure is currently under refinement but a preliminary analysis of the structure was carried out and the current refinement statistics are summarized in Table 2.6.

Table 2.6. Structure refinement (unfinished) statistics for ModA.

Refinement statistics	
Resolution range (Å)	44.75-3.0
R_{factor} **(%)**[a]	20.21
R_{free} **(%)**[b]	25.37
Average B-factor for all atoms (Å^2)	54.14
RMSD from ideal geometry	
Bond lengths (Å)	0.002
Bond angles (°)	0.423

[a] $R_{factor} = \sum ||F_{calc} - F_{obs}|| \,/\, \sum |F_{obs}| \times 100$, where F_{calc} and F_{obs} are the calculated and observed structure factor amplitudes, respectively.
[b] R_{free} is calculated for a randomly chosen 5% of the reflections.

2.2.5. Small-angle X-ray scattering of TupA and ModA

Samples of TupA and ModA were prepared in the presence and absence of tungstate or molybdate. For the measurements in the presence of the ligand, the protein was incubated with a 10-fold excess of sodium tungstate or sodium molybdate and equilibrated at 293 K for 25 min. The excess of the ligand was then removed by a PD MiniTrap G-25 column (GE Healthcare). The samples were concentrated above 12 mg/mL by ultrafiltration (Vivaspin® 500, Sartorius). Depending on the protein sample, different concentration ranges were tested (see Table 2.9 and 2.10 – section 2.3.3), between 32 and 0.6 mg/mL. All the samples were prepared in 5 mM Tris pH 7.6 and stored at 277 K until further used.

Data collection was performed in BM29 beamline at ESRF (France) and P12 beamline at Petra III (Germany) at 278 K and 280 K, respectively. Ten frames of 1 seconds each were collected, covering the range of momentum transfer $0.003 < s < 0.49$ Å^{-1} ($s = 4\pi\, sin\theta/\lambda$, where 2θ is the scattering angle) (Table 2.7). The primary data reduction and average of the collected frames was done automatically by the pipeline software EDNA[191] (for BM29, ESRF) or SASFLOW[81] (in P12, Petra III), both using the ATSAS package[85]. High and low concentration curves were merged to account for concentration effects such as interparticle interference using the program PRIMUS[192] from the ATSAS package. GNOM[88] was used to calculate the distance distribution function, $P(r)$, and determine the corresponding radius of gyration (R_g) and the maximum particle size (D_{max}) values. BUNCH[95] and SREFLEX[193] were used to generate and refine high-resolution hybrid models using the crystallographic structure of TupA and ModA. The scattering curves from the high-resolution models were calculated using CRYSOL [94].

The *ab initio* modeling programs DAMMIN[73] and DAMMIF[91] were employed for low-resolution shape generation, and 20 models were calculated in the slow mode, using default software settings. The program DAMAVER[75] was applied to superimpose individual structures and to determine the averaged and the most probable reconstruction.

The collected SAXS data and the generated high-resolution hybrid models for TupA are available at Small Angle Scattering Biological Data Bank (SASBDB) (entries: SASDBD9 (TupA apo model from BM29, ESRF), SASDBE9 (TupA apo model from P12, Petra III), SASDBF9 (TupA in presence of tungstate from BM29, ESRF), SASDBG9 (TupA in presence of tungstate from P12, Petra III) and SASDBH9 (TupA in presence of molybdate from BM29, ESRF)[194]. The models generated from ModA are under analysis and will be deposited later.

Table 2.7. Data collection parameters for the SAXS measurement of TupA and ModA.

Data collection parameters			
Beamline	P12 (Petra III, Germany)		BM29 (ESRF, France)
Protein samples	TupA TupA + WO_4^{2-} TupA + Mo_4^{2-}	ModA ModA + WO_4^{2-} ModA + Mo_4^{2-}	TupA TupA + WO_4^{2-}
Wavelength (Å)	1.24		0.99
s-range (Å^{-1})[a]	0.0027-0.45		0.003-0.49
Exposure time (s)	1		1
Temperature (K)	280		278

[a] Momentum transfer, $|s = \frac{4\pi sin\theta}{\lambda}|$

2.2.6. Urea-polyacrylamide gel electrophoresis

To performed the urea-polyacrylamide gel electrophoresis, TupA (at 50 µM) was incubated with a 10-fold excess of sodium tungstate or sodium molybdate (500 µM) and equilibrated at 293 K for 25 min. The excess ligand was removed by a PD MiniTrap G-25 column (GE Healthcare). The samples were analyzed using the Novex TBE-Urea polyacrylamide 6% gel and a XCell SureLock™ Mini-Cell Electrophoresis System (Invitrogen). A 1:1 mixture of protein and 2× Novex sample buffer was loaded into wells and the electrophoresis was carried out for 150 min at 180 V and 40 mA. To avoid the metal chelation, EDTA was removed from the electrophoresis solutions. This procedure was adapted from Mehtab *et al*[195], including the running and sample buffers recipes. The gel was stained with Coomassie blue solution.

2.2.7. Isothermal titration calorimetry of TupA and ModA

To perform the Isothermal titration calorimetry (ITC) assays, TupA (wild-type and mutants) and ModA were extensively dialyzed against 5 mM Tris-HCl pH 7.6 (reaction buffer). The same buffer was further used to prepare the ligand solutions, sodium tungstate dihydrate and sodium molybdate dihydrate, at 100 µM. The ITC experiments were performed using a VP-ITC calorimeter (MicroCal GE Healthcare), where 10 µM protein was added to the calorimetric cell and equilibrated at 303 K. After equilibration and baseline stabilization the protein solution was titrated with 20 or 23 injections of 10 µl of sodium tungstate or molybdate and the heat response recorded. Using the ORIGIN software package (Northampton, USA), the baseline was created, the peaks integrated and the area (µcal) under each peak obtained. After the manual adjustment of the baseline and integration details, the data fitted into a single binding site model and the stoichiometry of binding (n), the affinity constant (K_D) and the molar reaction enthalpy (ΔH) parameters were determined.

For the displacement titration assays, 100 µM of sodium molybdate (for TupA) or tungstate (for ModA) was added to the reaction buffer and the injections were performed with sodium tungstate or molybdate, respectively. The apparent binding affinity of the high-affinity ligand (L), K_{app}^{L}, can be determined by the binding affinity ($K_{a,X}$) and concentration of the weak ligand ([X]) – Equation 2.1 [196].

$$K_L^{app} = \frac{K_{a,L}}{1 + K_{a,X}[X]}$$

Equation 2.1.

2.3. Results and discussion

2.3.1. Structural characterization of TupA

2.3.1.1. Overall structure

Tungstate binding protein TupA is a monomer with 27 kDa and comprised by 251 residues (without purification tag or signal peptide). It has a globular and butterfly shape with approximate dimensions of 55.7 × 34.5 × 23.8 Å (Figure 2.10).

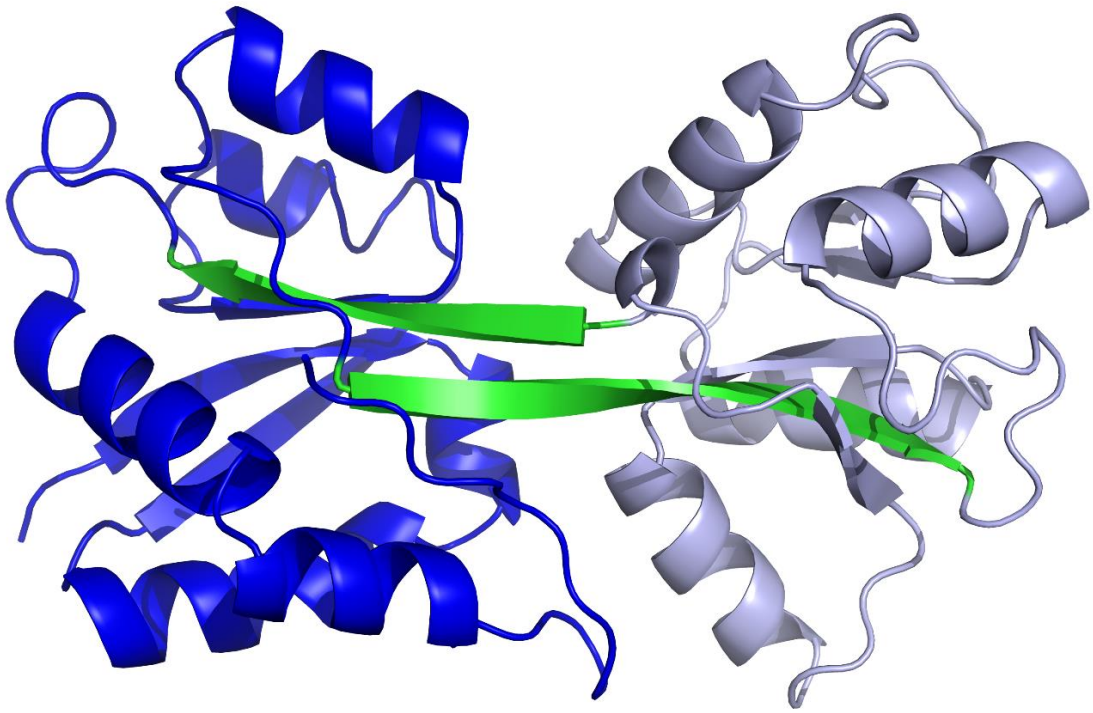


Figure 2.10. Cartoon representation of the *Da*G20 TupA tertiary structure. The different colors represent the two lobes, A in blue and B in light blue. In green is represented the hinge. Picture prepared with Pymol[197].

TupA has a mixed α/β bilobal structure where the two lobes of the protein are separated by a cleft. The lobe A (Figure 2.10, in blue) is formed by the residues A2-Y81 and Y201-E251; and the lobe B (light blue) by N82-Q200. Both the N- and C-terminal of the protein are in lobe A. As the polypeptide chain passes from the first lobe into the second and returns to the first, it creates two flexible amino-acid hinge segments that connect the lobes (Figure 2.10, in green). This architecture is a common feature of the substrate binding proteins (SBP) and was first described, in 1982, for the L-arabinose binding protein[198]. Later on, the same characteristic was found in ModA from *E. coli* [170] and *Archaeoglobus fulgidus* [115].

The central beta strand is hydrogen bonded to three beta strands of each of the lobes forming two 4-stranded beta sheets. Based on the connectivity of secondary structure elements and especially the topology of the beta-sheets core in the domains, the protein here described belongs to the Class II of the SBP – Fukami-Kobayashi[132] nomenclature system. TupA has $\beta_2\beta_1\beta_3\beta_n\beta_4$ as topology, with n representing the strand following the first cross-over from the N-terminal domain to the C-terminal domain – Figure 2.11. The larger segment of the TupA hinge is significantly longer, 11 amino acids (Figure 2.10) than the 4–5 amino acids hinges typically observed in several molybdate/tungstate binding proteins[123]. As such, a higher flexibility is possible between the open and closed conformation of TupA.

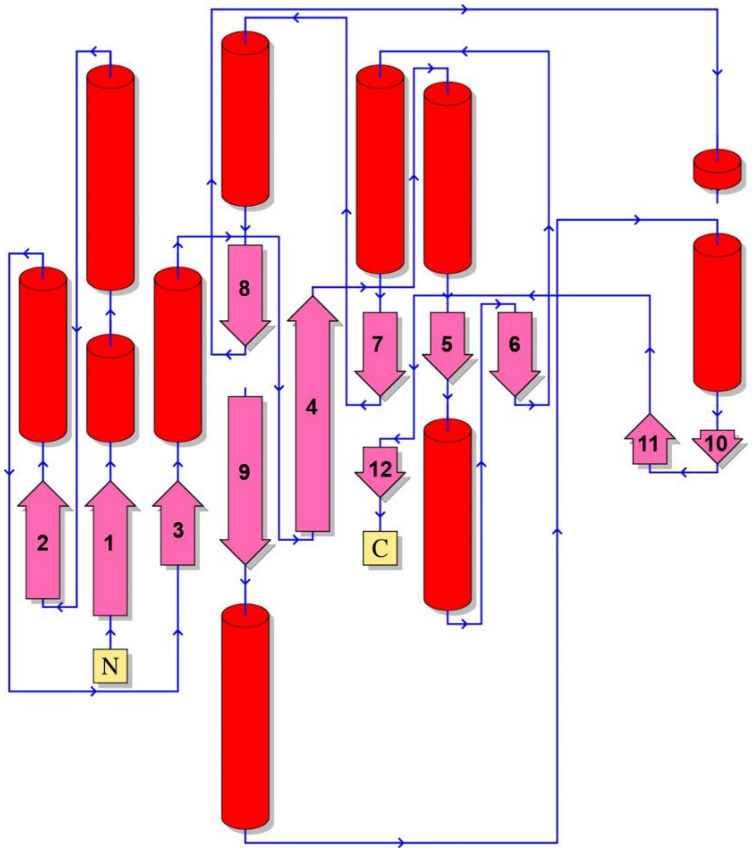


Figure 2.11. Topology diagram for TupA from *Desulfovibrio alaskensis G20*. The figure was prepared using the PDBsum server[199].

The TupA metal binding site is located within the central cleft formed by the two lobes. The two lobes share a very low sequence identity, about 5.2%, but their secondary structure elements and overall fold is very similar. In fact, superposition of the two lobes gives an RMSD of 3.0 Å (for 57 Cα) shows high resemblance between the two parts of the protein that holds the oxyanion – Figure 2.12. This could be associated with the fact that TupA is a monomeric protein suspected to interact with dimeric components of the ABC transporting system, like TupB. The structural similarities observed between the two lobes could be a result of an evolutionary strategy adopted to improve the interaction between the monomeric and dimeric, membrane associated,

components. This structural characteristic extends to other members of the TupABC family such as *G. sulfurreducens* TupA (UniProt Q749P2, PDB code 3lr1) (superposition of the two lobes gives a RMSD of 3.1 Å (for 37 Cα), but also to ModABC transporting family (*eg* ModA from *E. coli* [170], where an RMSD of 2.92 Å is obtained for superposition of 63 Cα) and WtpABC (*eg* ModA/WtpA from *Pyrococcus horikoshii* [169] with a RMSD of 3.05 Å from 65 Cα).

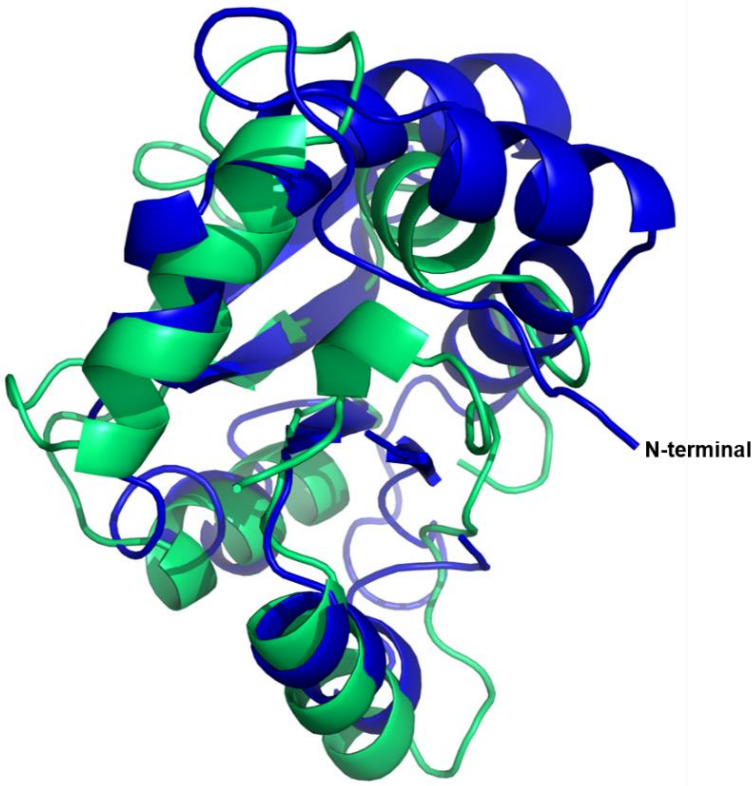


Figure 2.12. Superposition of the lobe A and B of TupA: lobe A (in blue) is formed by the residues A2-Y81 and Y201-E251; lobe B (light green) is composed by N82-Q200.

2.3.1.2. Comparison of *Da*G20 TupA with related structures

Several crystal structures of tungstate/molybdate binding proteins have been reported in the past few years. As mentioned before, the first was the ModA from *E. coli* at 1.75 Å (PDB code 1wod), in 1997 [170]. Later, in 2007, the crystal structure of the ModBC complex with ModA was solved at 3.1 Å resolution for *A. fulgidus* (PDB code 2onk) [115]. Both structures correspond to the holo-form in the presence of tungstate. Despite the low sequence identity between TupA and these two molybdate binding proteins (below 26%), the overall structure (RMSD over 3 Å) and the independent lobes are similar. The lobe A of TupA shares an RMSD of 1.82 Å (for 108 Cα superimposed) with ModA from *E. coli* and 2.03 Å (for 111 Cα) with ModA form *A. fulgidus*; from the superposition of lobes B results in an RMSD of 5.76 Å (34 Cα) for the *E. coli* structure and 2.32 Å (89 Cα) for *A. fulgidus*.

When the 3D model of TupA was used as a search model to query PDB, three bacterial proteins with over 45% sequence identity were obtained (Figure 2.13). These are the before mentioned *G. sulfurreducens* TupA, a protein with unknown function from *V. parahaemolyticus* (Q87PK2, PDB code 3muq), and a LysR, substrate binding domain from *W. succinogenes* (Q7M8V9, PDB code 3kn3), all obtained by the Protein Structure Initiative (PSI). In common, the orthologs contain the TTTS motif, other conserved residues for tungstate binding (like the R118) and the presence of a longer hinge (up 16 residues) between lobes, indicating that they should be classified as an SBP of the TupABC system.

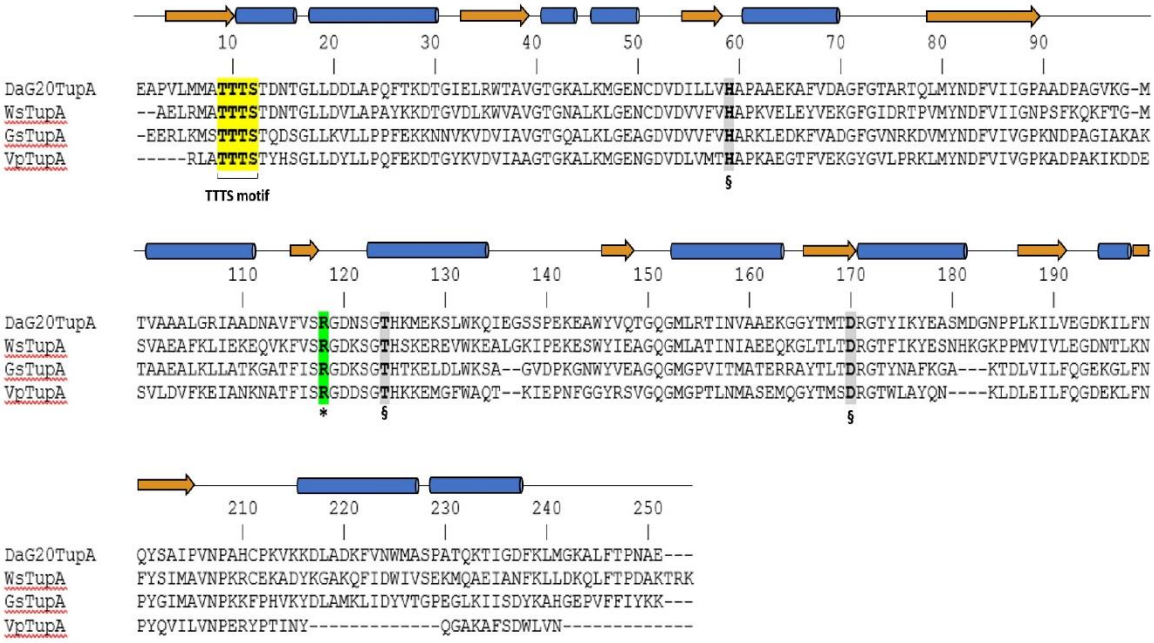


Figure 2.13. Multiple sequence alignment of mature TupA proteins from different organisms. DaG20TupA - TupA from *Desulfovibrio alaskensis* G20 (locus tag Dde_0234); WsTupA - TupA from *Wolinella succinogenes* (locus tag WS1370); GsTupA - TupA from *Geobacter sulfurreducens* (locus tag GSU2700) and VsTupA - TupA from *Vibrio parahaemolyticus* (locus tag VP1501). Important (*) and putative (§) residues involved in oxyanion binding.

The *Da*G20 TupA shares 45.6% sequence identity with *G. sulfurreducens* TupA, with conserved secondary structure elements, indicating that the two proteins are structurally very similar. However, the complete superposition of both proteins provides an unexpectedly high RMSD value (2.4 Å upon superposition of 208 Cα). A detailed comparison can be performed when considering the superposition of the independent lobes of the two proteins, yielding lower deviations (RMSD for the superposition of lobes A and B from the two proteins is 1.24 Å and 1.16 Å for 106 and 96 Cα, respectively). These values support a high structural similarity between *Da*G20 TupA and *G. sulfurreducens* TupA but also indicate some degree of flexibility of the lobes with respect to one another, as expected for the substrate binding protein of these transporters. *G. sulfurreducens* TupA was likely crystallized in the presence of an ionic form of tungsten, and the deposited model

reports a free W^{6+} accommodated in the metal binding cleft. In this TupA-W^{6+} holo form, the central cleft's volume (363.9 $Å^3$) is four times smaller than the apo form of *Da*G20 TupA (1480 $Å^3$). Unexpectedly, the cation is not coordinated to water molecules or protein residues. The water molecules surrounding the W^{6+} are separated by 3.23 - 3.82 Å and the closest residues are at 3.9 Å (the OG1 of T9 and NH1 of R118 (*G. sulfurreducens* TupA numbering) from lobes A and B, respectively). In *Da*G20 TupA these residues are at the same position although not superimposable with *G. sulfurreducens* TupA. The data suggest that conformational changes take place upon metal binding, where the protein in the holo form adopts a more compact conformation.

2.3.1.3. Oxyanion binding site

Most of the residues that form the binding cleft of ModA are essentially polar but poorly conserved among proteins from different organisms, as discussed for *E. coli* ModA [170]. In *Da*G20 TupA, the cleft is formed by positively charged residues as seen by the electrostatic surface potential calculations (Figure 2.14). The pronounced positive environment of the pocket must be an advantage to enable capture of the oxyanion, even when the extracellular concentration is low.

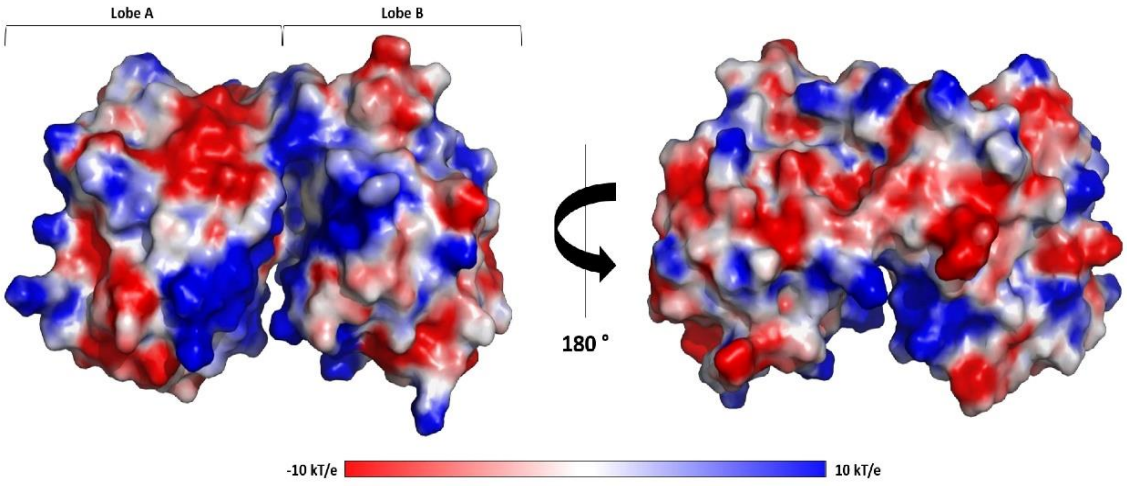


Figure 2.14. Electrostatic potentials of TupA surface. Electrostatic surface potentials were calculated using Pymol [197]. Surface potentials varies from −10.0 kT/e (red) to 10 kT.0 kT/e (blue).

As observed by Ledvina *et al*[200], more important than a complementary of electrostatic potential is the correct hydrogen-bonding capacity for the ligand binding. Several residues are likely to be involved in ligand binding, attracting, accommodating and delivering the oxyanion to the membrane component of the transporter system, TupB. The TTTS motif in lobe A is a conserved a.a. signature that indicates a selective binding of tungstate, with the serine (S12) pointing towards the metal binding site. In the same lobe, the H59 might be important for a selective protonation of tungstate, thus differentiating it from molybdate [171]. Opposite to these are R118,

T124 and D170, also likely involved in oxyanion interaction (Figure 2.15). These residues are highly conserved among other TupA proteins from *Desulfovibrio* species but also extending to proteobacteria, Green Non-Sulfur bacteria and even to Gram-positive bacteria such as Firmicutes. When searching for similar sequences with the TTTS motif and excluding the *Desulfovibrio* genus, over 500 sequences were found with more than 44% identity.

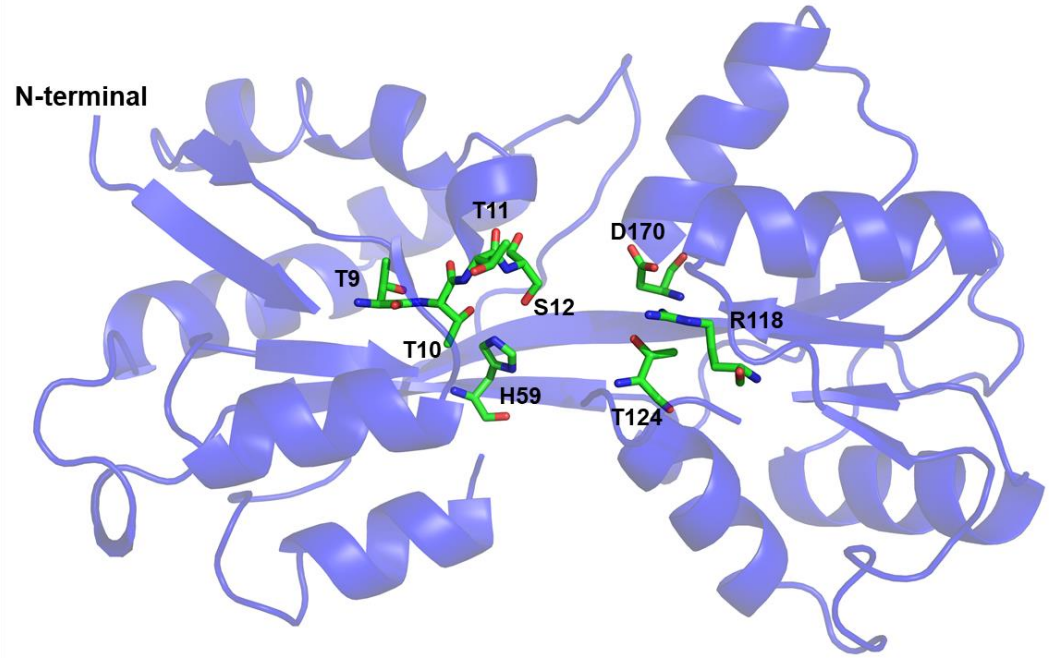


Figure 2.15. Cartoon representation of the *Da*G20 TupA 3D structure with the conserved residues involved in the metal binding site highlighted. The T9-11, H59, S12, R118, T124 and D170 are represented as sticks and color by element. Picture prepared with Pymol[197].

Moreover, one chloride anion, arising from the crystallization conditions (containing 0.2 M magnesium chloride), was found at 3.8 Å from R118, indicating the propensity of the pocket to attract negatively charged ions. Although phosphate buffer was used during protein purification this ion is not occupying the WO_4^{2-} binding site, in agreement with what is already known for this type of transporters and with our previous experimental data where we showed that *Da*G20 TupA is highly specific for WO_4^{2-} /MoO_4^{2-} and not for other oxyanions such as SO_4^{2-}, PO_4^{3-} and ClO_4^- (Figure A1, Appendix)[13]. To clarify the role of the highly conserved R118 in ligand binding, isothermal titration calorimetry was performed for three mutants, where this residue was mutated to a lysine (TupA_R118K), glutamine (TupA_R118Q) or glutamic acid (TupA_R118E), using site-specific mutagenesis. The results are described in section 2.3.4.1.

Several attempts were performed to determine the crystal structure of this protein in the presence of tungstate or molybdate, as mentioned in the previous section. Soaking and co-crystallization assays were carried out without success. The majority of the testing crystals had poor or non-existing diffraction. This could be due to the long and highly mobile hinge present in TupA,

allowing a more flexible and expanded movement between the apo and holo form. This feature could affect the packing of the protein leading to a poor X-ray diffraction.

2.3.2. Overall structure description of ModA

2.3.2.1. Overall structure and oxyanion binding site

One of the main objectives to study a tungstate and a molybdate binding protein from the same organism, *Da*G20, is to find structural evidences of how the proteins distinguish between two similar oxyanions.

The structure of the molybdate binding protein from *Da*G20, ModA, was solved by molecular replacement using the *A. vinelandii* ModA structure[139] as a query for the database. The model was refined to a maximum resolution of 3.0 Å, yielding an R_{factor} of 0.202 and R_{free} of 0.254 (Table 2.6, section 2.2.4). The structure is currently under refinement with the Ramachandran plot revealing that 96.3% of the residues were in the most favored regions and no outliers.

Like other molybdate binding proteins described in the literature, the ModA has a globular structure with dimensions of approximately 52 Å×30 Å×24 Å, presenting a 'Venus flytrap' shape typical of periplasmic SBP (Figure 2.16)[139,170,201].

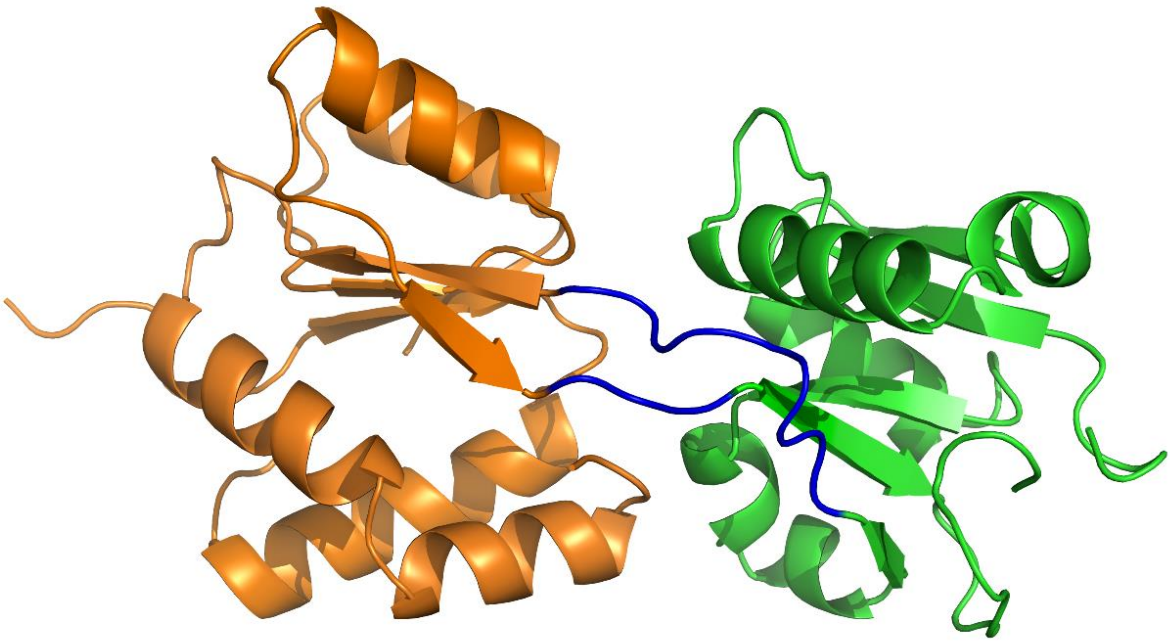


Figure 2.16. Cartoon representation of the *Da*G20 ModA 3D structure. The lobe A is represented in orange and the lobe B in green. In blue is the hinge between the two lobes. Picture prepared with Pymol[197].

The 26 kDa protein is formed, as expected, by two similar lobes (α/β sandwich), linked by a hinge, with the ligand-binding site located at the interface between the two lobes. There are two interdomain connections that make up the hinge. Consequently, lobe A is formed by two segments (1–79 and 182–223) and contains both N- and C- terminal, while lobe B is comprised of a continuous segment of the polypeptide chain (80-181).

Despite the low sequence identity between ModA and TupA (approximately 21%), the overall structure is quite similar with an RMSD of 2.46 Å. When analyzed separately the lobes present a lower RMSD at 1.65 Å (for 107 Cα) and 4.41 Å (for 35 Cα) for lobes A and B, respectively – Table 2.8. As other molybdate and tungstate proteins, ModA belongs to Class II of SBP[132] with a $\beta_2\beta_1\beta_3\beta_n\beta_4$ topology, with the two lobes connected by loops (Figure 2.17).

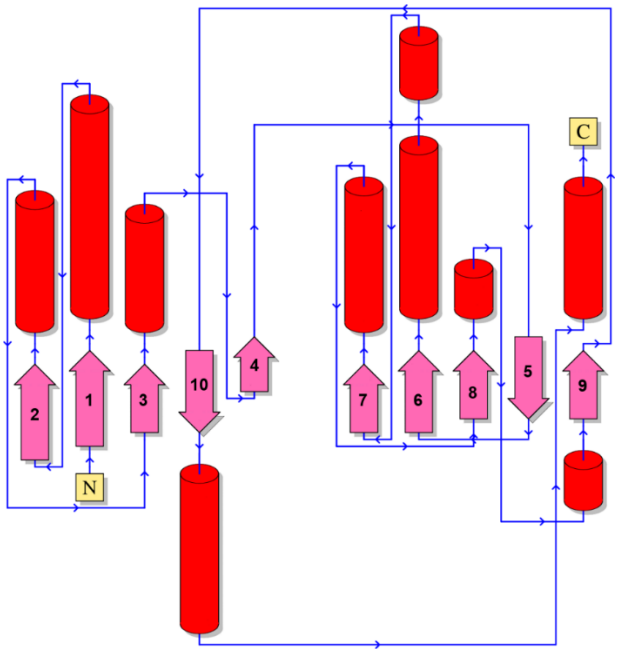


Figure 2.17. Topology diagram for ModA from *Desulfovibrio alaskensis* G20. The figure was prepared using the PDBsum server[199].

Curiously, the lobe A and B present a higher sequence identity, around 25%, when compared with TupA, which is about 5%. Based on 68 equivalent Cα atoms, the two lobes were superimposed with an RMSD of 3.34 Å.

The *Da*G20 ModA was compared with ModA from *A. vinelandii* and *A. fulgidus.* This protein shares 35% sequence identity with *A. vinelandii* ModA, with a similar overall structure. From the superposition of the two molybdate binding proteins results in an RMSD of 3.16 Å from 198 aligned residues. Again, a better superposition was obtained by considering the two lobes independently – Table 2.8. As described for TupA in the previous section (section 2.3.1.1.), this feature indicates the flexibility of the lobes with respect to one another as a consequence of the

hinge location, in the center of the structure. It is worth mentioning that the same result was obtained even when comparing ModA proteins from different organisms: *A. fulgidus* ModA shares 21% sequence identity with *Da*G20 ModA but from the superposition of lobes A and lobes B results a high RMSD when comparing the whole protein – Table 2.8.

Table 2.8. Comparison between *Da*G20 ModA with three related proteins. Root-mean-square deviation (RMSD) with the number of residues (in parenthesis) considered for the superposition of the 3D structures.

Proteins	**RMSD**			**Sequence identity (%)**
	Overall	Lobe A	Lobe B	
***A. vinelandii* ModA**	3.16 Å (198)	1.03 Å (117)	1.24 Å (90)	35
***A. fulgidus* ModA**	3.45 Å (186)	1.42 Å (110)	1.95 Å (90)	21
***D. alaskensis G20* TupA**	2.46 Å (169)	1.65 Å (107)	4.41 Å (35)	21

The ModA crystal structure reinforces the previous feature described for other ModA proteins[170,201] – where the binding cleft does not have a specific amino acid signature. Due to this fact, the identification of relevant residues involved in oxyanion coordination among the several ModA is not obvious by sequence analysis. Through structural interpretation and comparison with related structures in holo-form, it was possible to identify five residues in *Da*G20 ModA important for ion binding: N10, A56, I144, Y117 and Q187 (Figure 2.18). These residues mediate key protein-ligand interactions affecting both specificity and affinity of the protein. For molybdate binding proteins the most important are hydrogen bonds between the protein and the molybdate oxygens that contribute to the stabilization of the oxyanion negative charges[201].

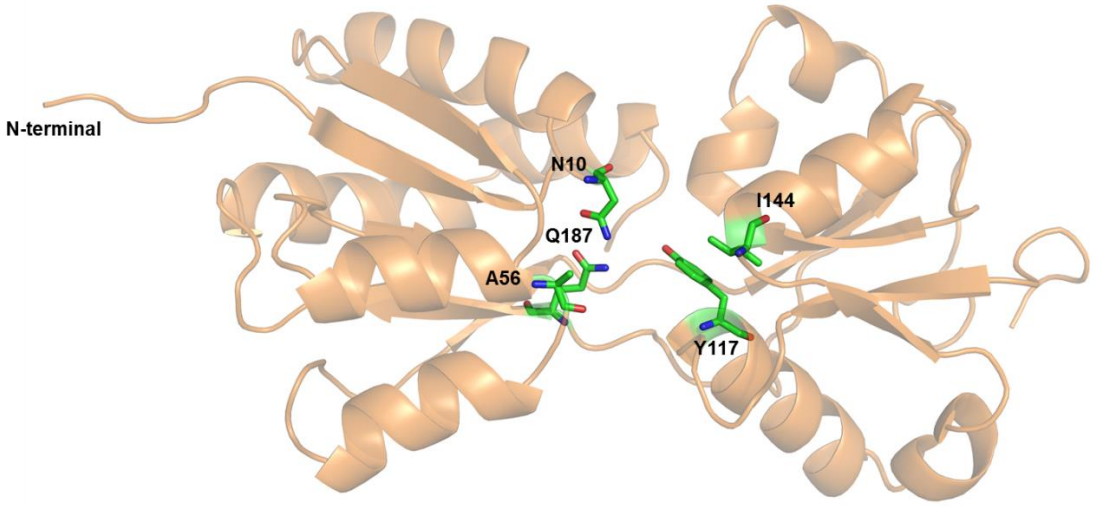


Figure 2.18. Cartoon representation of the ModA structure with the conserved residues involved in the oxyanion coordination highlighted. The N10, A56, Y117, I144 and Q187 are represented as sticks and color by element. Picture prepared with Pymol[197].

The N10, Y117 and Q187 residues are also conserved in *A. vinelandii*, but not in the other proteins considered for comparison. Although conserved in *A. vinelandii*, the Q187 is not part of the binding site. The A56 is conserved among the sequences under analysis (Figure 2.19). For *Da*G20 ModA, this residue is probably part of the ligand coordination since it is located in the center of the cleft and closer to the putative oxyanion binding site. Despite the variability of residues involved in the ligand binding, they are (or can be) all involved in the oxyanion coordination by hydrogen bonds donated by uncharged, polar protein groups. As observed for other ModA proteins that belong to bacteria domain, it is expected that in *Da*G20 ModA the oxygen atoms are tetrahedrally arranged around the metal center. This supports the previous observation that suggests distinct binding modes of the ModA/WtpA proteins from archaea and bacteria, with *A. fulgidus* or *Methanosarcina acetivorans* protein (and others) binding the oxyanion in an octahedral geometry[115,169,202]. Other residues present in the binding pocket of bacterial ModA (A8, A9, S35, S36, P116 and S153) are most likely enhancing the molybdate selectivity by contributing to the formation of an apolar environment [201].

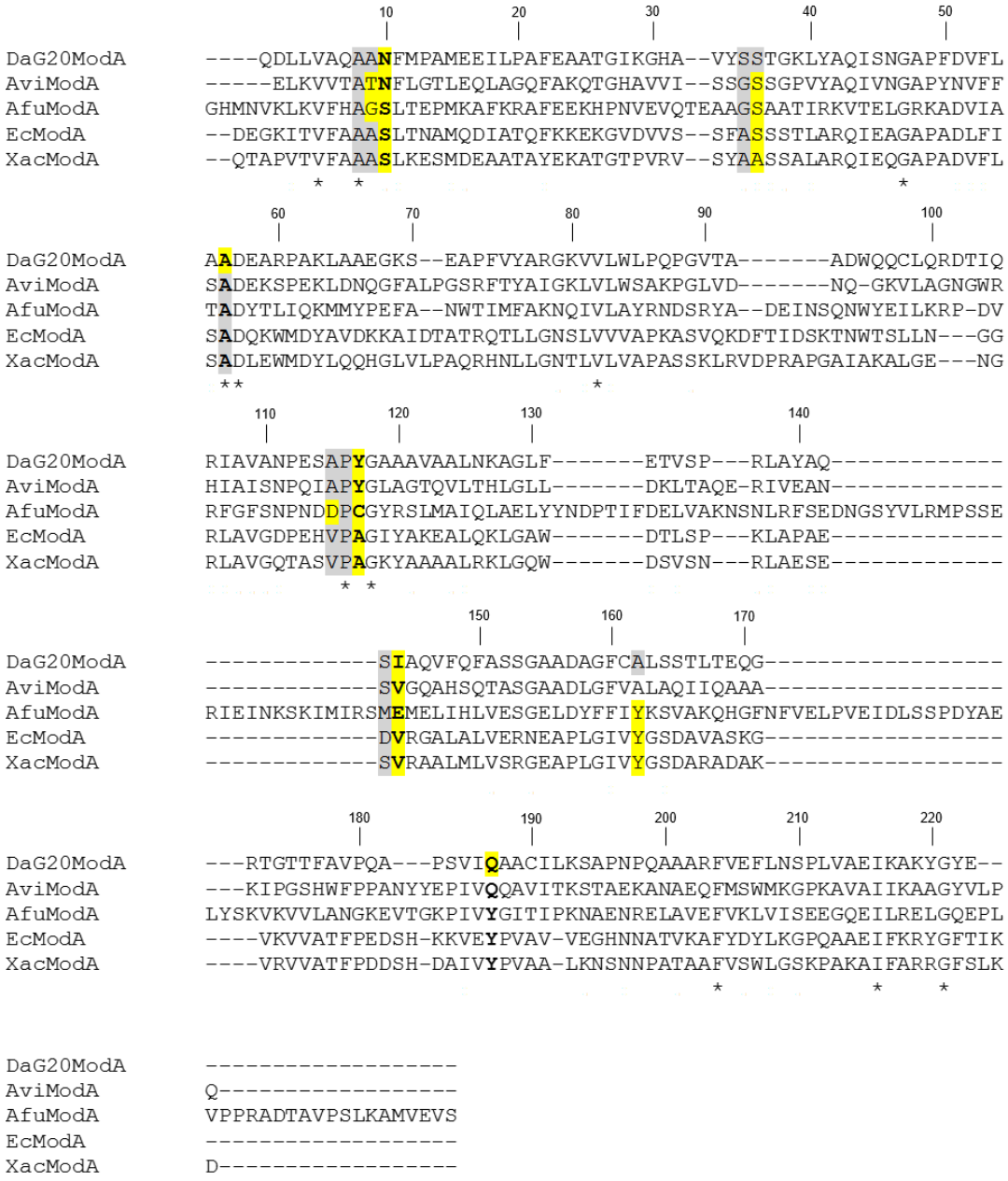


Figure 2.19. Comparison of the amino acid sequence of *Desulfovibrio alaskensis* G20 ModA with several orthologs. DaG20ModA - ModA from *Desulfovibrio alaskensis* G20 (UniProt code Q30VI5); AviModA - ModA from *Azotobacter vinelandii* (UniProt code Q7SIH2); AfuModA – ModA/WtpA from *Archaeoglobus fulgidus* (A0A101DFH9), EcModA - ModA from *Escherichia coli* (UniProt code A0A070SPW5) and XacModA – ModA from *Xanthomonas citri* (UniProt code A0A0D5UMP1). Highlighted in yellow are the residues involved in the oxyanion coordination and in gray the residues that form the binding pocket.

Comparing with the ModA binding site with TupA, the differences observed are remarkable. The ModA pocket is smaller, with an approximate volume of with 850 Å^3 (vs ~1500 Å^3 of TupA). Also, the identified metal binding residues are different. *Da*G20 ModA does not contain any positively charged residue at the binding site (Figure 2.20), contrarily to TupA that has an arginine and a histidine (R118 and H79), putatively important for binding. The TupA binding site is mainly

populated with polar residues, while the ModA pocket contains polar and apolar residues. The observed differences, the size and polarity, of the metal binding site could be the reason for the distinct binding affinities for molybdate and tungstate (more details in section 2.3.4.).

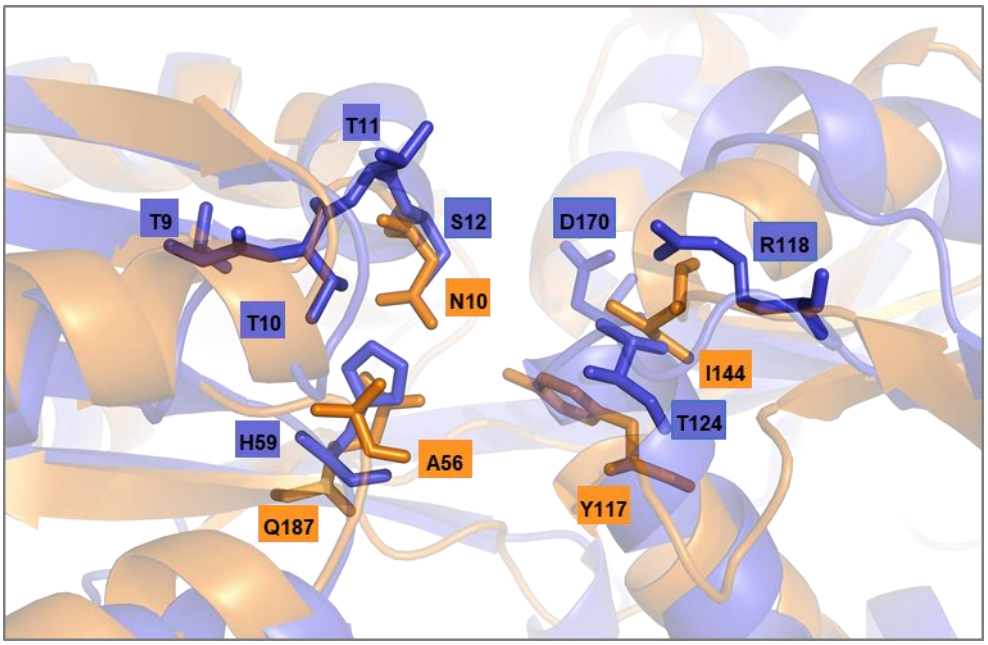


Figure 2.20. Binding site comparison between *Da*G20 TupA (blue) and ModA (orange). As sticks are represented the most important residues for the ligand binding. Picture prepared with Pymol[197].

2.3.2.2. Sequence homology and phylogenic analysis

The *modA* gene from *Da*G20 encodes a 256 residues protein, including a 33-amino acid signal peptide (MLSRLYALPARLLAPLCLPLCLTLCLLAAPAGA) for protein secretion across the bacterial cytoplasmic membrane. From the BLAST analysis against the nonredundant database (NCBI) results, resulted in the identification of 161 homolog proteins (E-value $<10^{-46}$). The protein with the highest sequence identity with ModA from *Da*G20 was the ModA from *Desulfovibrio longus* (*ca* 63%). The residues identified as important for coordination of oxyanion are conserved in *D. longus* (N10, A56, I144, Y117 and Q187, using *Da*G20 numbering). Nevertheless, the variability between the two is considerable, which is striking, since both play the same role and belong to the *Desulfovibrionaceae* family.

When the BLAST query was restricted to proteins available on PDB, seven hits were obtained for ModA proteins from *A. vinelandii*, *E. coli* and *X. citri*. As mentioned before, a significant difference exists in the amino acid sequence composition among the ModA orthologs. Considering ModA and the five orthologs with atomic structure, the phylogenetic analysis based on sequence similarity showed that these proteins can be grouped into three families (Figure 2.21).

Family 1, 2 and 3 are represented by the ModA from *Da*G20, *E. coli* and A. *fulgidus*, respectively. The most obvious differences in the oxyanion-binding residues occur between *Da*G20 and *E. coli*, with the two proteins sharing 26% of sequence identity (see Figure 2.19). In *Da*G20 ModA, the Y117 is a conserved residue throughout family 1 that, probably, coordinates the oxyanion through its main-chain NH. The conserved alanine (A115) is the equivalent residue in family 2. In families 2 and 3, Y170 establishes a hydrogen bond to the ligand via its side chain hydroxyl. However, in family 1 the equivalent residue (A162) it is only a 'passive' part of the binding cleft. Curiously, *Da*G20 ModA is the only member of family 1 where a glutamine (Q187) is probably involved in ligand binding.

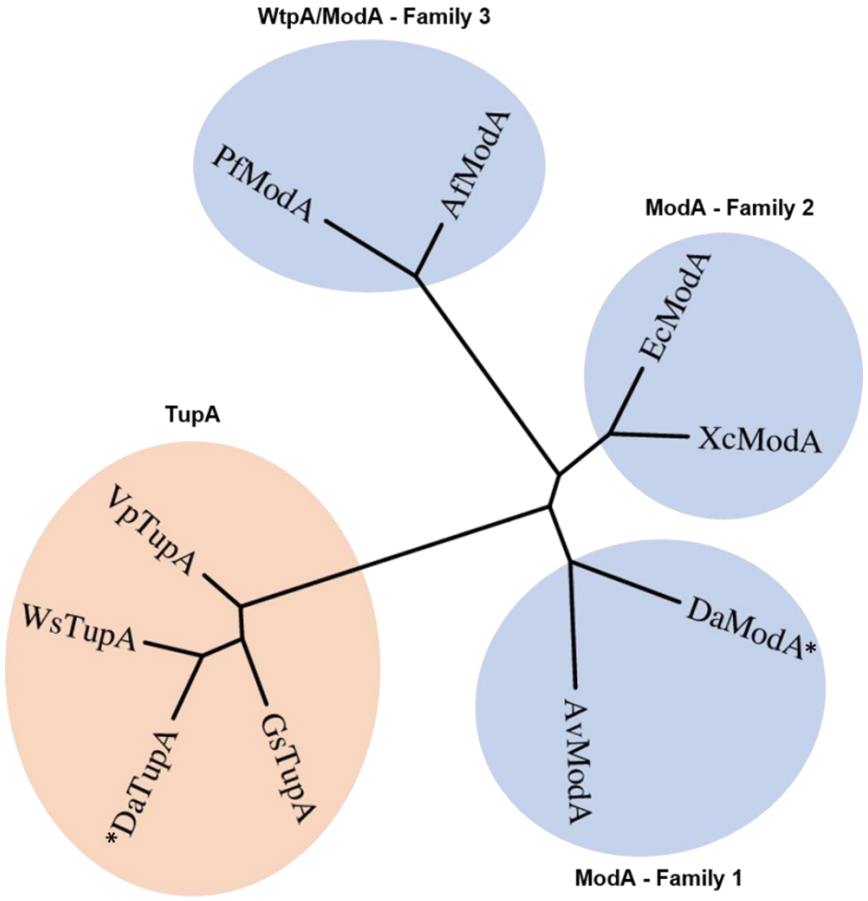


Figure 2.21. Phylogenetic analysis of *Desulfovibrio alaskensis* G20 ModA and orthologs. The unrooted tree generated with Phylogeny.fr[203] program. Family 1: DaModA - ModA from *Desulfovibrio alaskensis* G20 and AvModA - ModA from *Azotobacter vinelandii* (PDB code 1atg). Family 2: EcModA - ModA from *Escherichia coli* (PDB code 1amf) and XcModA – ModA from *Xanthomonas citri* (PDB code 2h5y. Family 3: AfModA – ModA/WtpA from *Archaeoglobus fulgidus* (PDB code 2onr) and PfModA – ModA from *Pyrococcus furiosus* (PDB code 3cg1). TupA: DaTupA - TupA from *Desulfovibrio alaskensis* G20 (PDB code 5my5), WsTupA - TupA from *Wolinella succinogenes* (PDB code 3kn3); GsTupA - TupA from *Geobacter sulfurreducens* (PDB code 3lr1) and VsTupA - TupA from *Vibrio parahaemolyticus* (PDB code 3muq). *Protein studied in this work.

Beyond the differences in the residues involved in oxyanion coordination (Figure 2.19) between the first two families and family 3, this last one is composed by two WtpA proteins from archaea

which bind the oxyanion in an octahedral coordination [169]. Homologues of the *wtpA* gene are not common in eubacteria evolution branch, being found only in three bacterial genomes (*Syntrophus aciditrophicus*, *Desulfotalea psychrophila*, and *Pelobacter carbinolicus*), and were not considered for phylogenetic analysis[169].

Sequences of three TupA proteins were added to the phylogenetic tree forming a separate group, family TupA (Figure 2.13 from section 2.3.1.2.). Although the two *Da*G20 proteins have the same physiological role, oxyanion uptake, they showed different anion affinities (see section 2.3.4. for more details). The phylogenetic tree illustrates that ModA and TupA are in different branches, suggesting that they may be derived from a common ancestral gene and have since diverged from the parent copy.

2.3.3. SAXS assays for protein envelope determination and ligand binding in solution

2.3.3.1. TupA scattering experiments

The results presented in this subsection were analyzed in collaboration with Prof. Dmitri Svergun and Dr. Alejandro Panjkovich from European Molecular Biology Laboratory-Hamburg Outstation, Germany.

To elucidate the possible conformational changes of *Da*G20 TupA upon ligand binding, synchrotron SAXS data were collected both in the presence and absence of tungstate and molybdate. For the SAXS measurements is necessary to guarantee that the reference buffer composition is the same as the sample to determine with accuracy the scattering derived from the protein. For that, the protein samples were passed through a PD MiniTrap G-25 column to remove the excess of ligand. The measurements were done at 278 K and 280K to avoid protein aggregation.

The scattering profiles for the apo and holo form of TupA are shown in Figure 2.22. From the SAXS scattering data, structural parameters were determined, including the radius of gyration (R_g), the maximum particle size (D_{max}) and the excluded volume of the hydrated particle (V_p) - (Table 2.9). The scattering data of the apo form indicate a monomeric globular protein with an R_g of 24.2 Å and a D_{max} of about 95 Å. In the presence of tungstate or molybdate (datasets TupA-WO_4^{2-} and TupA-MoO_4^{2-}, respectively), the overall shape of the protein remains globular but becomes more spherical and compact, leading to a decreasing on the R_g (~23 Å) and D_{max} (~90 Å) values. To compare the datasets obtained in the presence of tungstate and molybdate, the tool CorMap [204] from the ATSA software was applied. Importantly, the datasets are very similar, revealing no statistically significant difference (C=12, *p*-value=0.12) between them and indicating that in solution the protein adopts the same conformation upon binding both oxyanions.

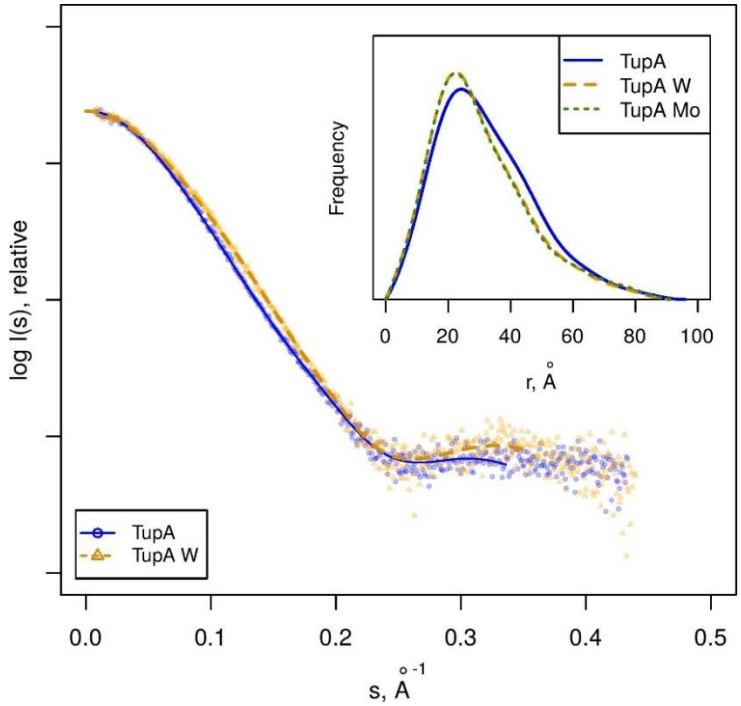


Figure 2.22. SAXS scattering data (points) and GNOM fits (lines) for TupA in the absence (TupA) and presence of tungstate (TupA W). The data collected in presence of molybdate was omitted from the main plot for clarity as it matches the TupA W data up to noise. Inset: distance distribution functions, $P(r)$, for the different conditions measured.

Table 2.9. Structural parameters obtained by SAXS for TupA protein in the presence or absence of oxyanion. $\chi 2$ values correspond to discrepancies between models and experimental data, the lowest $\chi 2$ value per dataset is highlighted.

	TupA	**TupA***	**TupA W**	**TupA W***	**TupA Mo**
Concentration range (mg/mL)	1.0-8.5	0.8-6.0	1.0-6.0	0.8-12.0	1.0-6.0
Structural parameters					
R_g, Å [from $p(r)$]	25.5 ± 0.3	25.4 ± 0.3	24.1 ± 0.2	23.8 ± 0.2	24.4 ± 0.2
R_g, Å (from Guinier)	24.2 ±0.4	24.0 ± 0.4	23.0 ± 0.3	22.8 ± 0.3	22.9 ± 0.3
D_{max}, Å	96 ± 3	95 ± 3	90 ± 3	89 ± 3	92 ± 3
Porod volume V_p, Å^3 x 10^3	51.4 ± 3.5	48.5 ± 3.3	45.7 ± 3.1	44.4 ± 3.0	45.8 ± 3.1
Molecular mass determination (kDa)					
From protein sequence			29.6		
From $I(0)$[a]	29 ± 2	32 ± 2	26 ± 2	29 ± 2	26 ± 2
From Porod volume[b]	31 ± 2	29 ± 2	28 ± 2	27 ± 2	28 ± 2
Model vs. data discrepancy ($\chi 2$)					
Holo-form model	4.0	2.4	**0.9**	**2.2**	**0.9**
Crystal structure	7.8	5.8	1.8	11.6	1.8
Apo-form model	**1.0**	**0.7**	2.0	11.9	1.7

[a] Using BSA as a reference. [b] Calculated by $MM = V_p/1.7$

Looking at the inset of Figure 2.22, the comparison of the distance distribution functions, $P(r)$, between holo and apo forms shows a more compact conformation for the holo form. This is in agreement with the idea that the protein adopts a more compact structure upon binding to the metal ion, which is consistent with the 'Venus flytrap' model seen for bacterial-type periplasmic binding proteins[104].

The N-terminal region contains fusion sequences from the expression vector (MA**HHHHHH**VDDDDKMLEVLFQGP) which was expected to remain unstructured since is not visible in the TupA crystal structure at 1.4 Å resolution. This is also supported by the SAXS data where a slowly decaying long inter-distance tail in the $P(r)$ function at large distances is observed (Figure 2.22, inset).

The TupA crystal structure was used as a starting point to create models for both the holo-form and apo-form states. The 23 N-terminal residues that are missing in the crystal structure were calculated using BUNCH[95]. This program is used to determine a tridimensional domain structure based on scattering data. Based on *G. sulfurreducens* TupA-W^{6+} adduct, a tungstate group was added to the expected binding site of the BUNCH model, generating a theoretical holo-form model. The theoretical scattering curves calculated using this holo-form model of TupA are in very good agreement with the SAXS measurements of the protein in the presence of tungstate or molybdate (discrepancy, chi-square, $\chi^2 = 0.9$) (Figure 2.23, Table 2.9). The apo-form model was created by refining the initial BUNCH model with the SREFLEX program[193], which is used to improve the agreement between high-resolution models with experimental SAXS data. This approach revealed a slight opening of the lobes (RMSD of 1.5 Å for 274 Cα atoms) yielding an excellent agreement ($\chi^2 = 1.0$) to the TupA SAXS data measured in the absence of tungstate or molybdate.

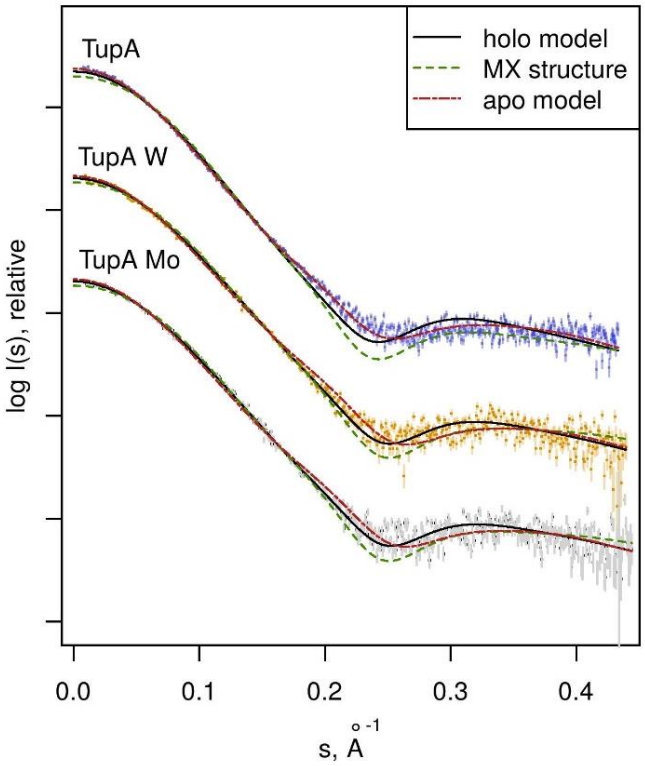


Figure 2.23. SAXS scattering data (points) for the three experimental conditions, TupA in the absence (TupA) and the presence of tungstate (TupA W) or molybdate (TupA Mo). Each dataset was scaled for display purposes. For each experimental curve, CRYSOL fits are displayed for the crystallographic structure reported in this work (MX) and the holo- and apo-form models generated thereof. The best fit for TupA is the apo-form model ($\chi^2 = 1.0$), while the best fit for both TupA W and TupA Mo is the holo-form model ($\chi^2 = 0.9$).

The optimum holo-form model yielding the best χ^2 to the experimental data and the smallest RMSD to the original structure is presented in Figure 2.24. The conformational transition between the apo and holo form is represented by the gray arrows and the modeled N-terminal by the small blue spheres. The SAXS data corroborates that *Da*G20 TupA is a flexible protein that adopts a loose conformation in the free form and upon binding to molybdate or tungstate, switches to a more compact fold.

Since TupA adopts the 'Venus Fly-trap' mechanism upon ligand binding, in the presence of the ligand occurs a stabilization of the closed conformation. To complement the SAXS data and study the impact of the metal binding in the protein stability, a urea-polyacrylamide gel electrophoresis was carried out for TupA in the presence/absence of tungstate and molybdate. In Figure 2.24 (inset), the gel shows that, upon tungstate binding, the protein migrates further than in the absence of metal or in the presence of molybdate. This indicates that TupA adopts a more compact conformation that is likely to increase stability under a 7 M urea gradient, in agreement with our SAXS analysis (Figure 2.24, inset).

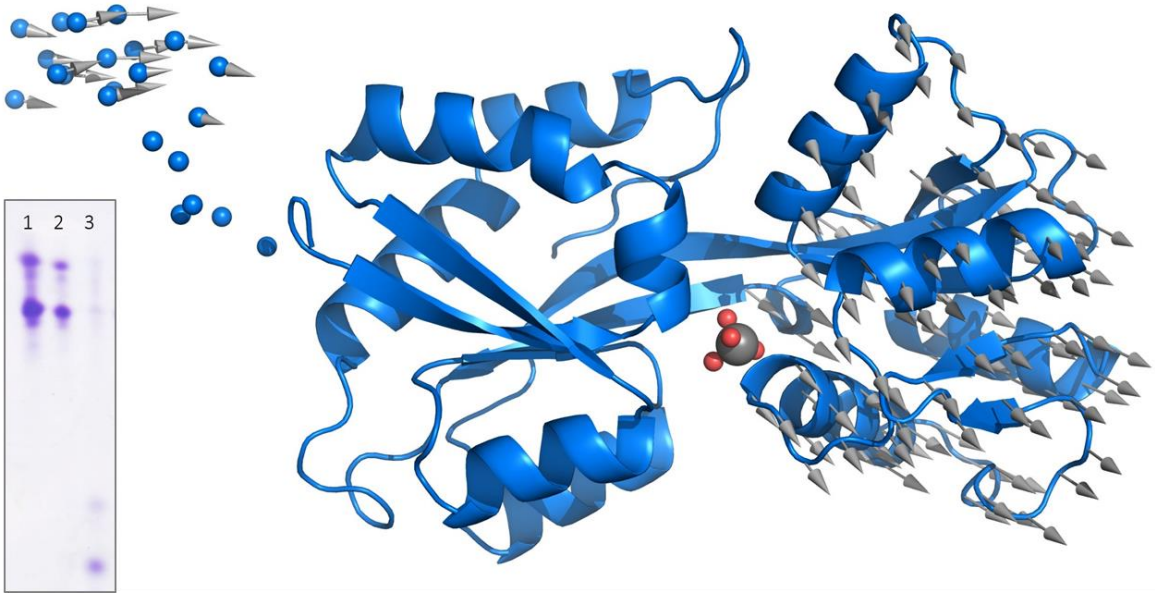


Figure 2.24. Cartoon representation of the tridimensional coordinates for the holo-form hybrid model of TupA. The N-terminal section modeled with BUNCH is shown as small spheres. The large gray sphere in the center corresponds to the tungstate group (small red spheres represent O atoms) modeled by homology with PDB entry 3cfz and 3lr1. Vectors have been drawn connecting Cα from the holo-form model to the apo-form model generated by SREFLEX, after superposition of lobe A, to display the 'opening' conformational transition. Upon optimal superposition including all Cα (274), the RMSD is 1.5 Å.
Inset: Urea-polyacrylamide gel electrophoresis of 1) TupA, 2) TupA+MoO_4, 3) TupA+WO_4. The samples in presence of ligand were first passed through a size exclusion PD-10 minitrap G-25 columns to eliminate the excess.

2.3.3.2. ModA SAXS analysis and comparison with TupA

Due to the similarity of the measurement setup and overall structure of ModA with TupA (RMSD 2.40 Å from 169 Cα), the same approach described in the previous sub-section 2.3.3.1 was applied. From the scattering profiles for ModA in the absence/presence of tungstate and molybdate, the structural parameters R_g, D_{max} and V_p were determined – Figure 2.25 and Table 2.10.

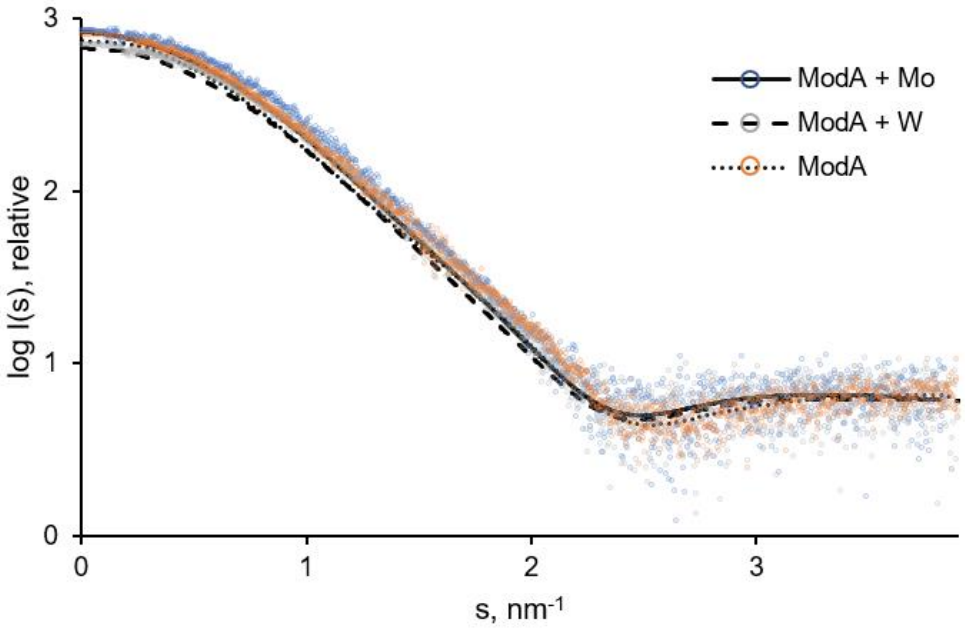


Figure 2.25. SAXS scattering data (points) and GNOM fits (lines) for ModA in the absence (ModA) and presence of tungstate (ModA + W) or molybdate (ModA + Mo).

Table 2.10. Structural parameters obtained by SAXS for ModA protein in the presence or absence of oxyanion. χ^2 values correspond to discrepancies between models and experimental data. The best χ^2 value per dataset is highlighted.

	ModA	**ModA + Mo**	**ModA + W**
Concentration range (mg/mL)	1.4 – 11.0	1.0 - 6.0	1.0 - 6.0
Structural parameters			
R_g, nm [from $p(r)$]	2.06	1.89	1.84
R_g, Å (from Guinier)	1.93	1.97	1.95
D_{max} (nm)	6.44	6.17	6.03
Porod volume V_p (nm^3)	38.23	38.40	36.47
Molecular mass determination (kDa)			
From protein sequence		26.1	
From Porod volume[a]	22.5	22.6	21.5
Model vs. data discrepancy (χ^2)			
Holo-form model	1.35	**0.82**	**0.77**
Crystal structure	0.62	0.66	0.72
Apo-form model	**1.10**	0.76	0.75

[a]Calculated by $MM = V_p/1.7$

The scattering profile of the apo-ModA indicates a monomeric globular protein with an R_g of 2.06 nm Å and a D_{max} of 6.44 nm. In the presence of molybdate or tungstate the protein remains the same overall shape but, a small decrease in the R_g and D_{max} values are observed, indicating a

holo-form protein with a more compact conformation. Curiously, for ModA the differences between the parameters of the apo and holo forms are not so pronounce as observed for TupA (section 2.3.3.1, Table 2.9). This could indicate a smaller conformational change upon ligand binding in the case of ModA. In fact, looking at crystal structure of both proteins (section 2.3.1.1 and 2.3.2.1), we observed that the hinge in ModA (connects the two lobes) is smaller than in TupA. This feature has impact in the flexibility of the protein which is, probably, lower in ModA justifying the small differences observed between the apo and holo-form.

By the analysis of the distance distribution functions, $P(r)$, in the presence and absence of oxyanion (Figure 2.26), a more compact conformation for the holo-form can be observed, supporting the idea that ModA adopts the 'Venus flytrap' model for the ligand binding. Although both $P(r)$ curves indicate a more compact envelop, the distance distribution functions in presence of tungstate or molybdate are not superposed – suggesting that the binding to the different oxyanions have different impact on the overall structures. Since other techniques (see sections 2.3.2 and 2.3.4) supports the protein binding of these two oxyanions, this difference should be addressed by X-ray crystallography.

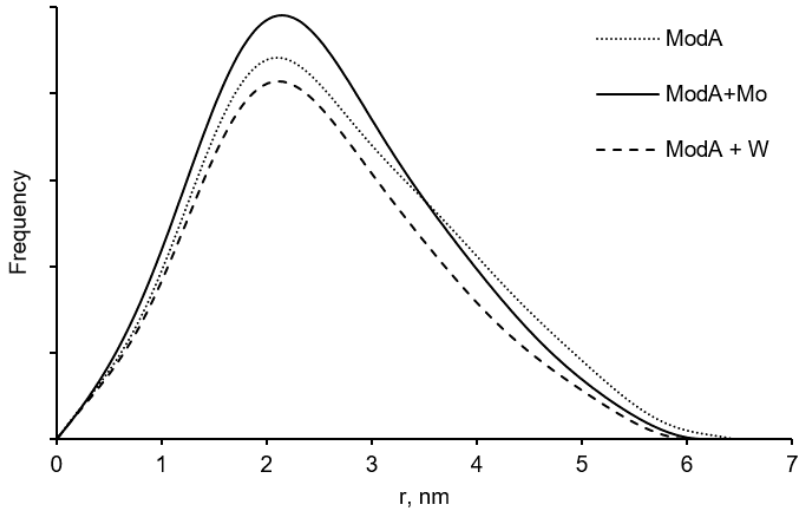


Figure 2.26. Distance distribution functions, $P(r)$, for ModA in absence (ModA) or presence of tungstate (ModA + W) or molybdate (ModA + Mo).

Overall, the observed differences between the ModA and TupA SAXS parameters were expected. They are derived from the slight difference in the molecular weight (ModA – 26.9 kDa; TupA – 29.1 kDa) between the two proteins and, especially, from the size of the hinge. As mentioned before, the TupA has a larger hinge than ModA and this feature have impact in the behavior of the protein in solution. The parameter R_g and D_{max} reflects these difference, with TupA having a larger diameter and radius of gyration than ModA – Table 2.9 and Table 2.10.

The protein ModA also contains an N-terminal region that contains a purification tag with 23 residues (MA**HHHHHH**VDDDDKMLEVLFQGP), which, again, is not complete in the crystal

structure of the apo-form. In order to have a more detailed comparison between the SAXS data and the crystal structure, BUNCH[95] was used to calculate the contribution of the extra residues based on scattering data (similar to what has been performed for TupA analysis). The generated model was then refined using SREFLEX[193], creating the apo-form model. Based on the crystal structure of *A. vinelandii* ModA[139] (PDB code 1atg), a tungstate group was added to the expected binding site of the BUNCH model, generating a theoretical holo-form model.

Using CRYSOL[94], the theoretical scattering profile from apo- and holo-form models were calculated and compared with the experimental scattering curve from SAXS. The results are summarized in Table 2.10 and Figure 2.27. Looking at the χ^2 values obtained for the apo-form, we observe that from the BUNCH[95] software results a model with an excellent agreement with the SAXS data (χ^2 of 1.10), comparing with the result obtained using crystal structure directly (χ^2 of 0.62).

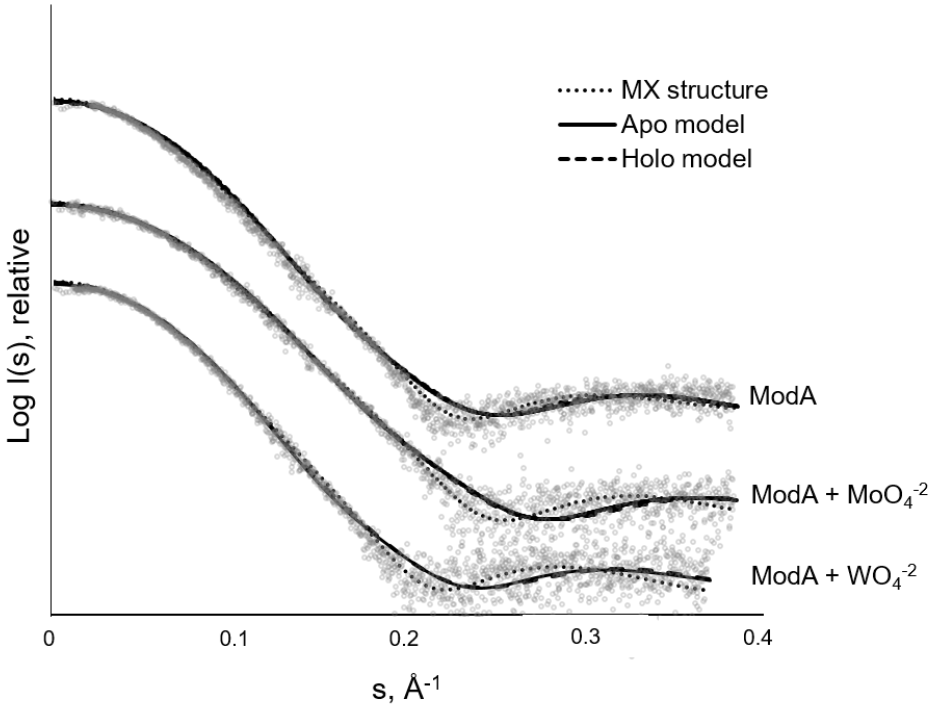


Figure 2.27. SAXS scattering data (points) for the three experimental conditions, ModA in the absence (ModA) and the presence of tungstate (ModA + WO_4^{2-}) or molybdate (ModA + MoO_4^{2-}). Each dataset was scaled for display purposes. For each experimental curve, CRYSOL fits are displayed for the crystallographic structure reported in this work (MX) and the holo- and apo-form models generated thereof. The best fit for ModA is the apo-form model (χ^2 = 1.10), while the best fit for both MoO_4^{2-} and ModA WO_4^{-2} is the holo-form model (χ^2 ~ 0.8).

The theoretical holo-form model is also in agreement with the scattering data collected in the presence of molybdate (χ^2 0.82) and tungstate (χ^2 0.77), supporting the hypothesis that ModA adopts a more compact conformation upon ligand binding. As shown by the primary data analysis,

it is evident that the shift between the apo and holo-form does not lead to very large conformational change since the scattering data also fits the apo-form model.

Using the program DAMMIF[91], the *ab initio* shape of ModA was determined by simulated annealing using a single-phase dummy atom model. After refinement with DAMMIN[73], the 20 models obtained were average with DAMAVER[75]. This program suite aligns *ab initio* models, selects the most probable and builds an average model. All calculated models have a normalized spatial discrepancy (NSD) less than 0.5, and tending to 0, which indicates a systematical similarity between them.

Finally, the most probable model derived from the scattering data was superimposed with the crystallographic structure of ModA using SUPCOMB[205] (Figure 2.28). From superposition results an NSD of 1.09, which indicates a symmetrical difference between the models. In fact, the results are not surprising due to the different behavior of the protein molecules in solution and in the crystal state.

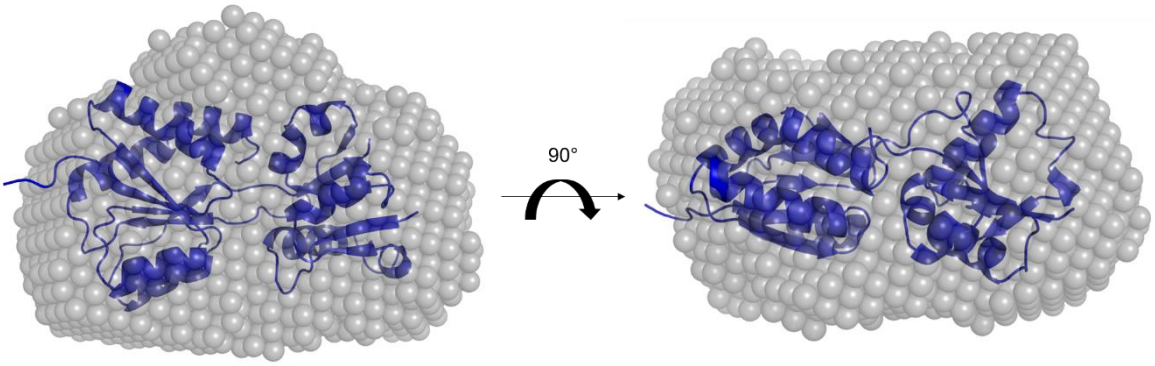


Figure 2.28. Superposition of the *ab initio* envelope of ModA with the cartoon representation of the crystal structure. Transparent beads are the most typical reconstruction from twenty DAMMIF runs.

Overall, the results described in this section show the importance of using different techniques to characterized biological systems and processes. Due to the difficulties to determine the crystal structures of the tungstate/molybdate binding proteins from *Da*G20 in the presence of the oxyanion, SAXS helped to confirm that TupA and ModA adopt the 'Venus flytrap' model.

2.3.4. Metal binding affinity characterization

Isothermal titration calorimetry (ITC) is a widely used biophysical technique for studying the formation or dissociation of molecular complexes. It is a direct measurement of the heat generated or absorbed when molecules interact. ITC has been proven to be a sensitive method to determine affinity constants for tungstate and molybdate-binding proteins, TupA and ModA, in the nanomolar and subnanomolar ranges[158,171,172]. Since it is expected a single binding site, from a sigmoidal titration curve, the stoichiometry (n), dissociation constant (K_D), and change in enthalpy (ΔH) can be directly measured.

When using this technique is important to assure the macromolecule saturation with a ligand, where no more binding occurs and only heat of dilution is observed. For that, a 10-fold excess of ligand to titrate the protein was used.

2.3.4.1. TupA wild-type and mutants

The first assays were performed with TupA in the presence of tungstate or molybdate. The ITC titration curves and binding isotherms are represented in Figure 2.29. The binding isotherms were created by plotting the integrated heat peaks as a function of the molar ratio of the ligand. These titration curves (top of Figure 2.29) suffer adjustments to compensate some experimental features, such as the fluctuation of cabinet temperature of the ITC chamber for each point and the normalization of baseline temperature with the corresponding buffer controls. They also acted as a quality parameter, demonstrating the reliability of the data.

The observed behavior from the analysis of Figure 2.29A is consistent with an exothermic process, at 303 K, with a single binding site model. However, the extremely high affinity of the protein for tungstate (K_D of 0.5 ± 0.4 nM) resulted in a very steep binding curve, which hampers the determination of K_D (Table 2.11). For a reliable determination of the K_D from a single titration, the values should be typically limited to the range of 100 μM > K_D > 1 nM [206].

A standard protocol to overcome this problem is the displacement titration assay [196,206], that will allow the calculation of the correct affinities. Since TupA also binds molybdate, this methodology was applied. In this setup, the tungstate was titrated into a solution containing the TupA saturated with molybdate. The affinity for the tungstate decrease because it has to displace the molybdate from the binding site in order to be able to interact with the protein but, using the equation 2.1 described in section 2.2.7, the parameters were deduced.

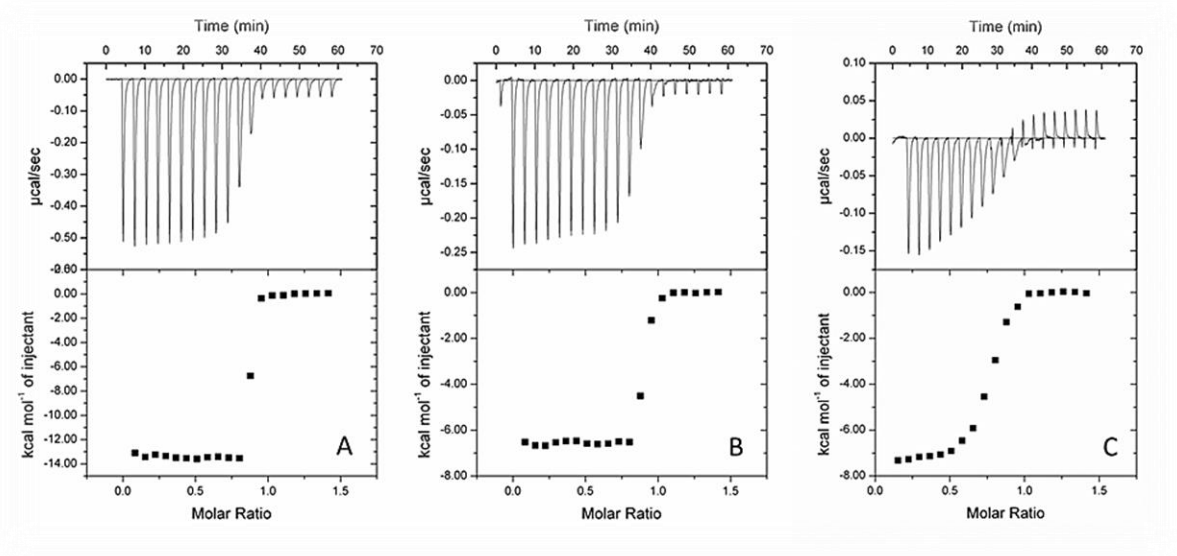


Figure 2.29. Isothermal titration calorimetry of ligand binding to TupA. TupA (10 µM) was titrated with injections of 100 µM tungstate (A) and 100 µM molybdate (B); (C) Displacement titration of 10 µM TupA incubated with 0.5 nM molybdate, with injections of 100 µM tungstate. Data were fitted with ORIGIN software. The raw ITC data are shown in the top graphs.

Table 2.11. Data for the ITC analysis of oxyanion binding to TupA protein at 303 K. In each case, 10 µM protein was used for the titrations.

	Ligand	n	K_A (M^{-1})	K_D (nM)	ΔH (kcal mol^{-1})
TupA	WO_4^{2-}	0.842 ± 0.001	$2.2x10^9 \pm 1.9x10^9$	0.5 ± 0.4	-13.5 ± 0.005
	MoO_4^{2-}	0.868 ± 0.002	$1.6x10^8 \pm 2.3x10^7$	6.1 ± 0.86	-6.6 ± 0.003
TupA + 0.5 mM MoO_4^{2-}	WO_4^{2-}	0.845 ± 0.003	$1.6x10^{11} \pm 5.7x10^8$	$6.3x10^{-3} \pm 2.2x10^{-5}$	-14.6 ± 0.04
TupA + 0.5 mM WO_4^{2-}	**MoO_4^{2-}**	**No displacement**			

n = measured stoichiometry of binding.

The results (Table 2.11) shows that TupA binds exothermically both oxyanions with a stoichiometry of one-mole oxyanion per mole of protein, deduced from the heat release upon the addition of tungstate or molybdate to the protein solution (Figure 2.29). From the direct titration of sodium molybdate against TupA results in an exothermic binding reaction with a K_D value of 6.1 ± 0.9 nM. Looking at the value of ΔH obtain for this oxyanion (approximately -6.6 kcal/mol of injectant), it reveals that the molybdate binding is less favorable when compared with tungstate. In contrast, the binding of tungstate to TupA release even more heat, with an increase of the ΔH to, approximately, -14 kcal/mol of injectant (Table 2.11). A displacement titration of the molybdate-saturated protein with tungstate clearly showed that the protein favors the binding of tungstate

(K_D 6.30 ± 0.02 pM), even when the binding site is occupied with a molybdate molecule (0.5 mM of molybate). This difference of at least three orders of magnitude observe between tungstate and molybdate, indicates the ability of TupA to discriminate very sharply between these two similar oxyanion. The described dissociation constants is in agreement with what has been observed for the putative *C. jejuni* 1540 TupA, with this protein also binding more tightly tungstate (K_D 1.0 ± 0.2 pM) than molybdate (K_D 50 ± 10 nM) [172]. Curiosity the K_D of TupA for tungstate is, approximately, 1000 times higher than the K_D value described for the *E. acidaminophilum* TupA (1 nM, obtained from a competitive assay)[171].

To understand the relevance of the R118 in the oxyanion binding affinity, three mutants were produced by site-directed mutagenesis. The R118 was changed to a glutamine, glutamic acid or lysine. The reasoning behind this mutagenesis is that the substituting amino acids have a different charge (glutamine and glutamic acid) and size (for lysine). The same methodology described for TupA wild-type was used to characterize the binding of the mutants. The results are described in Figure 2.30 and Table 2.12.

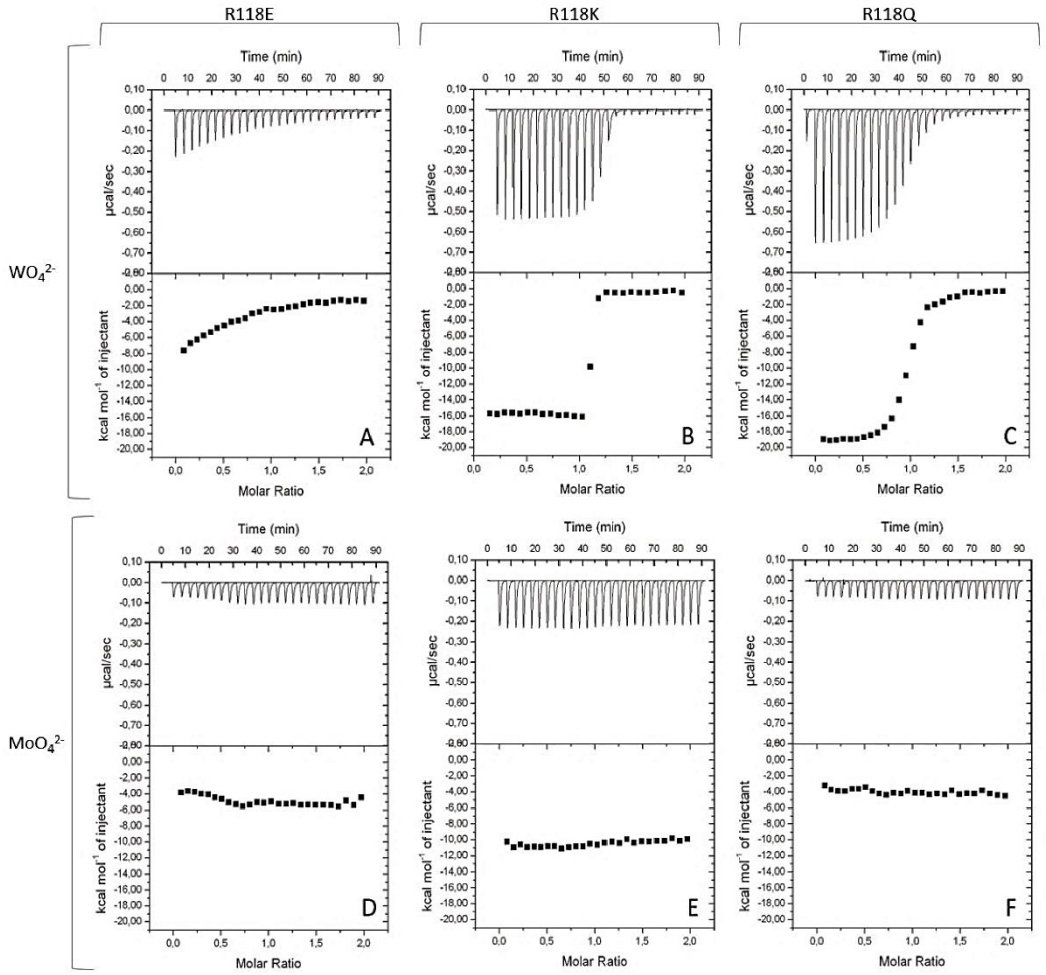


Figure 2.30. Isothermal titration calorimetry of ligand binding to TupA mutants. 10 µM of the mutant R118E (A, D), R118K (B, E) or R118Q (C, F) was titrated with injections of 100 µM of Na_2WO_4 (A, B, C) and 100 µM of Na_2MoO_4 (D, E, F). Data were fitted with ORIGIN software. The raw ITC data are shown in the top graphs.

Table 2.12. Data for the ITC analysis of tungstate binding to TupA mutants at 303 K. In each case, 10 µM protein was used for the titrations.

TupA mutants	n	K_A (M^{-1})	K_D (nM)	ΔH (kcal mol^{-1})
R118K	1.080 ± 0.003	(6 ± 2) × 10^8	1.8 ± 0.8	-15.8 ± 0.1
R118Q	0.950 ± 0.003	(0.111± 0.006) × 10^8	90 ± 50	-19.40 ± 0.08
R118E	No binding			

n = measured stoichiometry of binding.

By the analysis of the ITC titration curves and binding isotherms, it is clear that the substitution of the positive side chain by an uncharge (glutamine), negatively charged (glutamic acid) or an amino acid with a shorter side chain (lysine) abolishes the interaction between the TupA mutants and the molybdate oxyanion – see injection profile and binding isotherms in Figure 2.30 D, E, F.

For the tungstate, the substitution of the arginine by glutamine (K_D 90 nM) greatly decreases the affinity of interaction by four orders magnitude - see Figure 2.30 C and Table 2.12. The substitution of the R118 for a glutamic acid also abolishes the interaction with tungstate – Figure 2.29 A. This reveals that a dramatic change in charge in this position has severe consequences in the protein activity, confirming the relevance of this conserved residue. However, if the positive charge is kept in position 118 by replacing the arginine by a lysine, the results are remarkably different. Analysis of the ITC curves of R118K titrated with tungstate (Figure 2.30 B) shows a very steep binding curve, indicating a strong interaction. The obtained curve hampers the correct determination of K_D, as verified previously for the wild-type. The same strategy adopted before, displacement titration method, cannot be applied considering that this mutant lost the ability to bind molybdate. Nevertheless, the results clearly show that R118K mutant has an extremely high affinity and selectivity for tungstate, with an erroneous K_D of 1.8 ± 0.8 nm – value obtained by fitting the model to the steep binding curve. This contrasts with what has been previously observed for *E. acidaminophilum* TupA, where the authors reported that the mutagenesis of an extending positively charged residue of arginine to lysine strongly diminishes the specific binding of tungstate [156,171].

An interesting question is how the protein accomplishes this specificity at the molecular level. A discrimination by size seems unlikely because both anions are nearly identical in size. TupA might be able to take advantage of the difference in the pKa value of tungstate (4.60) regarding molybdate (3.87)[207], that have implications in the hydrogen bond strength and in the affinity towards the oxyanion[166]. These ITC results support this hypothesis showing the importance of the R118 in oxyanion binding affinity/coordination.

In the future, ITC assays should be performed for other TupA mutants, namely the H59. As mentioned before, this residue can be crucial in the oxoanion ambiguity present in the tungstate/molybdate binding proteins.

2.3.4.2. ModA and comparison with TupA

To characterize the oxyanion affinity of ModA, ITC measurements were performed in the presence of molybdate and tungstate. The same methodology described before was used and results are presented in Figure 2.31 and Table 2.13.

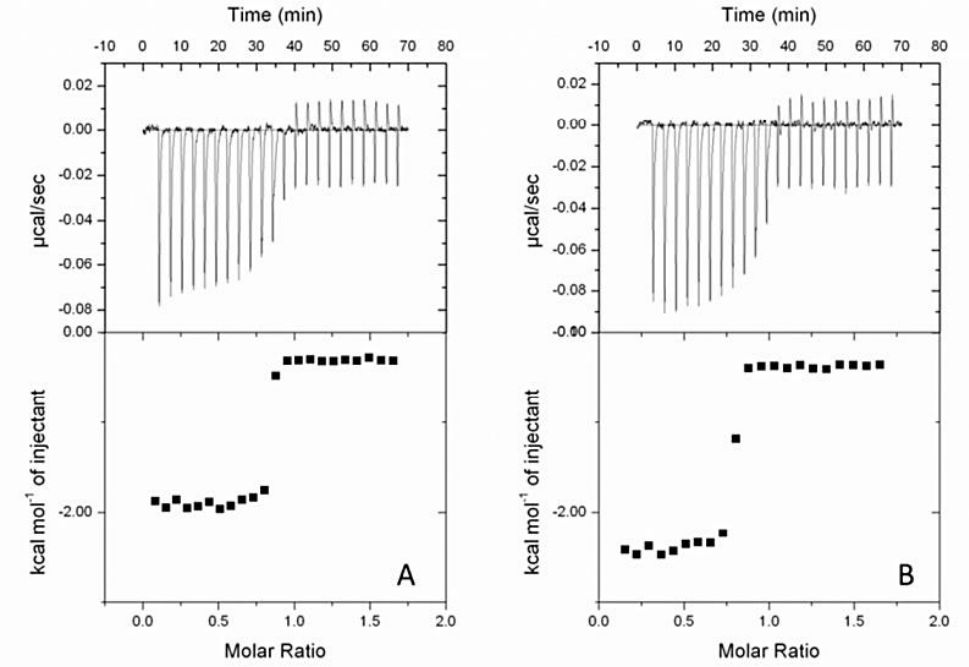


Figure 2.31. Isothermal titration calorimetry of ligand binding for ModA. ModA (10 µM) was titrated with injections of 100 µM tungstate (A) and 100 µM molybdate (B). The raw ITC data are shown in the top graphs.

Table 2.13. Data for the ITC analysis of oxyanion binding to ModA at 303 K. In each case, 10 µM protein was used for the titrations

	Ligand	n	K_A (M^{-1})	K_D (nM)	ΔH (kcal mol^{-1})
ModA	WO_4^{2-}	0.820 ± 0.01	$9x10^7 \pm 8.0x10^7$	12 ± 10	-1.9 ± 0.01
	MoO_4^{2-}	0.781 ± 0.002	$2x10^7 \pm 1x10^7$	43 ± 25	-2.7 ± 0.01
ModA + 0.1 mM MoO_4^{2-}	WO_4^{2-}	No displacement			
ModA + 0.1 mM WO_4^{2-}	MoO_4^{2-}	No displacement			

n = measured stoichiometry of binding.

The ITC titration curves and binding isotherms show that ModA binds exothermically molybdate and tungstate ions. At 303 K, the protein has a high affinity towards tungstate and molybdate (12 and 43 nM, respectively), resulting in a very steep binding curve, which hampers the

determination of K_D. Using the same strategy adopted for TupA, displacement titration assays were performed saturating the ModA with 0.1 mM of ligand. These experiments reveal that ModA binds both oxyanions with very high and similar affinities, being unable to replace one oxyanion for the other in a displacement titration assay. The same behavior was observed for other ModA, such as ModA from *C. jejuni* (Cj0303, with a K_D for both ligands of 4–8 nM) [172] and *E. coli* (with a K_D for the two oxyanion of 0.13-0.17 nM)[168]. As described for other ModA proteins (such as ModA from *E. coli* [167,168]), due to the incapability to distinguish between the two similar oxyanions, these are probably, equally taken up by the cell. Supporting this fact, the crystal structures of ModA with bound molybdate or tungstate have been shown to be identical, excluding any redox reaction of the oxyanions upon binding[170]. Based on this, the discrimination between both oxyanions must then occur along the molybdenum cofactor biosynthesis, probably at some step dependent on the redox state of the metal[167].

These results contrast with what has been observed for *Da*G20 TupA, where this protein is able to distinguish between the two oxyanions, binding tungstate with a higher affinity than molybdate (6.3 pM vs 6.1 nM, respectively). As discussed previously, the observed differences in the metal binding affinity can result from the different size and polarity of the binding pocket, determined in the crystal structures of the two proteins.

After transported into the cell, molybdate and tungstate will be incorporated in the biosynthesis of cofactors that are crucial for the activity of multiple enzymes. In Chapter 3, a more detailed information can be found on the next steps.

This study contributed with a structural analysis of a tungstate- and molybdate-binding proteins, showing the main features and the residues involved in oxyanion coordination. By complementing with ITC analysis it is clear that a TupA binds tungstate with higher affinity than molybdate, while ModA binds the two oxyanions with the same affinity. It is yet not clear how the protein distinguishes between the two ligands and, for this, mutagenesis studies should be carried out.

Chapter 3

Escherichia coli Periplasmic aldehyde oxidoreductase (PaoABC) and its chaperone (PaoD)

Part of the work described in this chapter was the subject of three publications:

- Correia MA*****, **Otrelo-Cardoso AR***, Schwuchow V, Clauss KGVS, Haumann M, Romão MJ, Leimkühler S, Santos-Silva T. The *Escherichia coli* periplasmic aldehyde oxidoreductase is an exceptional member of the xanthine oxidase family of molybdoenzymes. ACS Chem Biol. **2016**; 11(10): 2923–35. *****These authors contributed equally to this work.

- **Otrelo-Cardoso AR**, Correia MA, Schwuchow V, Svergun DI, Romão MJ, Leimkühler S, Santos-Silva T. Structural data on the periplasmic aldehyde oxidoreductase PaoABC from *Escherichia coli*: SAXS and preliminary X-ray crystallography analysis. Int J Mol Sci. **2014**; 15(2): 2223-36.

- **Otrelo-Cardoso AR**, Schwuchow V, Rodrigues D, Cabrita EJ, Leimkühler S, Romão MJ, Santos-Silva T. Biochemical, stabilization and crystallization studies on a molecular chaperone (PaoD) involved in the maturation of molybdoenzymes. PlosOne. **2014**; 9(1): e87295.

The work described in this Chapter results from collaboration with Prof. Silke Leimkühler from Institut für Biologie und Biochemie, Universität Potsdam, Germany.

3.1. Introduction

Mo-dependent enzymes (molybdoenzymes) contain in most cases a molybdenum atom coordinated to a dithiolene group of a tricyclic pyranopterin monophosphate named molybdopterin (MPT). This forms the molybdenum cofactor (MPT or Moco)[208,209]. The cofactor allows the right positioning of catalytic metal at the enzyme's active site, modulates its redox behavior, and mediates the electron transfer to or from the Mo atom[12].

3.1.1. Molybdenum cofactor

Once in the cell, Mo has to be attached to its cofactor scaffold, forming the Moco and gaining biological activity. In 1971, Nason *et al* demonstrated the existence of a molybdenum-containing component shared by known molybdoenzymes at the time[210]. Twenty years later, Rajagopalan and Johnson presented the first model for the Moco biosynthesis in *Escherichia coli*[16]. In all kingdoms of life, Moco is synthesized by a conserved biosynthetic pathway that can be divided into four steps (Figure 3.1): (1) formation of the cyclic pyranopterin monophosphate (cPMP); (2) conversion of cPMP into MPT by introduction of two sulfur atoms; (3) adenylation of MPT; and finally, (4) the insertion of molybdenum to form Moco. In prokaryotes, a fifth step may be present and includes the modification of Moco by the addition of a nucleotide, thus forming bis-molybdopterin guanine dinucleotide (bis-MGD) or molybdopterin cytosine dinucleotide (MCD) cofactors [211,212].

The pyranopterin is the only pterin known to be substituted with a four-carbon side chain. The biosynthesis of the cPMP (or precursor Z) starts with the 5'-GTP (guanosine-5'-triphosphate), which is converted into cPMP by two enzymes (MoaA and MoaC in *E. coli*). MoaA is a member of the superfamily of S-adenosyl-methionine (SAM)-dependent radical enzymes, being a homodimer with two [4Fe-4S] clusters in each monomer. This protein forms an (8S)-3′,8-cyclo-7,8-dihydroguanosine 5′triphosphate (3′,8-cH_2GTP) intermediate that is converted by MoaC to cPMP. GTP labeling and NMR studies in *E. coli* demonstrated that each carbon of the ribose and guanine are incorporated into the first intermediate of Moco biosynthesis. The cPMP is the most stable intermediate of Moco biosynthesis, with an estimated half-life of several hours at low pH. For this reason, cPMP is one of the most viable options in the treatment of the Moco-free human sulfite oxidase deficiency[213]. In 2013, Clinch *et al* reported that the chemical synthesis of a tricyclic pyranopterin intermediate can be successfully converted *in vitro* into a functional Moco, since it contains all four chiral centers present in the cPMP[214]. This study represented the first synthetic

route for a biologically active derivative of Moco, providing the basis for the future treatment of Human Moco deficiency (MoCD) type A patients[215].

In the second step, the MPT dithiolate is formed by the incorporation of two sulfur atoms at the C1′ and C2′ positions of the cPMP backbone. This reaction is catalyzed by the heterotetrameric complex, MPT synthase, formed by two small (MoaD) and two large (MoaE) subunits that form an $(\alpha\beta)_2$ heterodimer. The complex is formed by dimerization of two large subunits forming two clearly separated active sites that act independently during the reaction[216]. After MPT synthase transfers the two sulfur atoms to cPMP, the MoaD has to be re-sulfurated in order to regenerate the enzyme for the next cycle [217]. In bacteria, MoeB is the protein responsible for the reactivation of MPT synthase. This protein activates the MoaD by adenylation of the C-terminal glycine followed the addition of a sulfur atom from the pyridoxal-dependent IscS sulfurtransferase [217,218]. Recently, a new protein involved in the Moco biosynthesis in *E. coli* was identified, TusA. TusA interacts with IscS for the thiomodification of tRNAs, but also play a role in Moco biosynthesis. This implies that IscS does not directly interact with MoaD, but rather transfers the sulfur to TusA in a sulfur relay system, which then transfers the sulfur to MoaD [7,219]. Experimental evidences may suggest that TusA balances the distribution of IscS-labilized sulfur among the several sulfur-requiring metabolic pathways existent in *E. coli* [7].

The third and four steps, results in the adenylation of MPT and insertion of a molybdenum atom. In *E. coli*, this step is catalyzed by two independent proteins (MogA and MoeA), whereas during the evolution of higher organisms, these two proteins were fused to a single two-domain molybdenum insertase (Cnx1 in plants and gephyrin in mammals)[220]. First, the MPT needs to be activated by the MogA protein through adenylation – formation of the MPT-AMP intermediate. Then, the MPT-AMP if transferred to MoeA which cleaves the adenylate from MPT and catalyzes the insertion of molybdenum into the dithiolene group of MPT, forming Moco or Mo-MPT[221].

While the described steps of Moco biosynthesis are conserved in all kingdoms of life, a further modification of Moco only exists in prokaryotes. This modification includes the addition of a nucleotide (guanine or cytosine) to the phosphate group of Mo-MPT, forming MGD or MCD, respectively[7]. The biosynthesis of the bis-MGD occurs in two steps and requires Mo-MPT, the protein MobA and Mg-GTP. First, the bis-Mo-MPT intermediate from two Mo-MPT is formed by MobA. In the second step, two GMP moieties from GTP are added to the C4' phosphate of each MPT, forming the bis-MDG cofactor[222]. Contrarily, the formation of the MCD is a one step reaction catalyzed by the *E. coli* MocA protein. MocA acts as a molybdopterin CTP transferase and covalently bond Mo-MPT and CMP with the concomitant release of the β- and γ-phosphates of CTP as pyrophosphate[223].

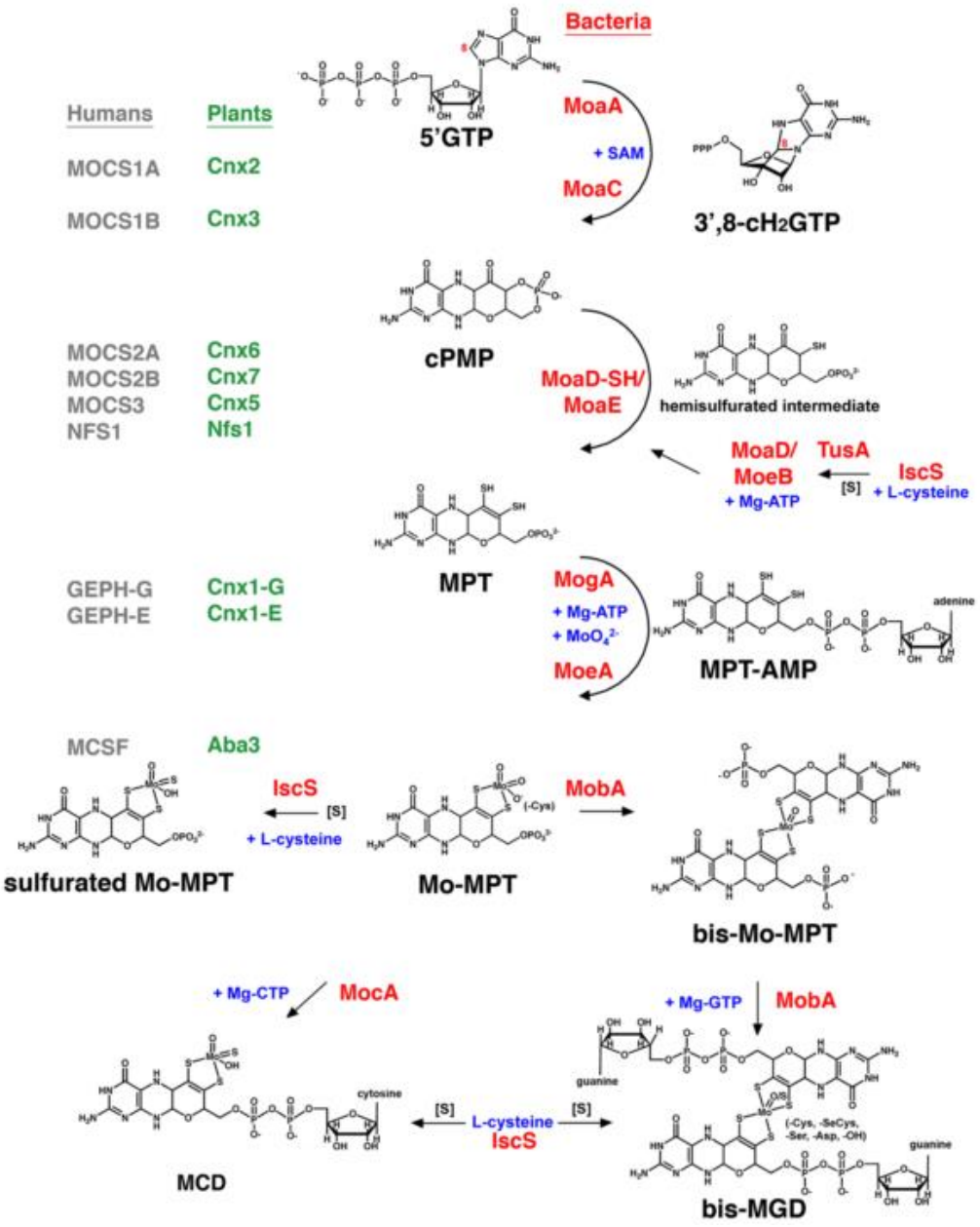


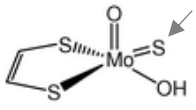
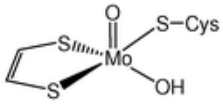
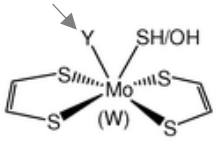
Figure 3.1. Biosynthesis of the molybdenum cofactor. The scheme contains the proteins involved in the reactions for bacteria (in red), plants (green) and humans (gray). Additional co-substrates required for the reactions are colored in blue. Adapted from[7].

3.1.2. The molybdoenzymes families

Molybdoenzymes are found in almost all branches of life. Typically, the Mo-dependent enzymes catalyze an oxo-transfer reaction coupled to electron-transfer between the substrate and other cofactors, such as [2Fe-2S] and/or [4Fe-4S] and/or hemes[15]. More than 60 different molybdoenzymes have been identified. The vast majority of these are present in prokaryotes while in animals and plants only four and five different enzymes, respectively, have been characterized [224]. Based on X-ray crystallography data, spectroscopy and biochemical analysis,

the molybdoenzymes have been divided into three families: the xanthine oxidase (XO) family, the sulfite oxidase (SO) (and assimilatory nitrate reductases) family and the DMSO reductase family (Table 3.1). Enzymes of the DMSO reductase are present exclusively in prokaryotes, while the XO and SO families are found also in eukaryotes.

Table 3.1. Schematic representation of the molybdenum cofactor in the different families of molybdoenzymes. Adapted from[15,224].

Xanthine oxidase family	Sulfite oxidase family	DMSO reductase family
		
(or O, Se, -S-Cu-S-Cys)		(Y: -S-Cys, -Se-Cys, -O-Ser, -O-Asp, -OH)

The **SO family** contains important enzymes for the sulfur metabolism in plants, animals, and bacteria, but also the assimilatory eukaryotic nitrate reductase, responsible for the reduction of nitrate to nitrite [225]. These enzymes are thought to be true oxygen atom transferases, with either a $LMo^{VI}O_2(S{-}Cys)$ as oxygen atom donor or $LMo^{IV}O(OH)(S{-}Cys)$ as an acceptor in the active site[224]. Here, the distinctive feature is to have the polypeptide chain directly coordinated to the molybdenum, through a cysteine residue[15,224].

X-ray crystal structures of the SO from chicken SO[226], *Arabidopsis thaliana* SO[227], and *Starkeya novella* sulfite dehydrogenases (SDH)[228] show nearly identical square pyramidal coordination around the Mo atom, even though the overall structures of the proteins and the presence of additional cofactors vary [225].

In eukaryotes, SO is located in the mitochondrial intermembrane space and is responsible for the oxidative degradation of the methionine and cysteine (sulfur-containing amino acids), using cytochrome c as the physiologic electron acceptor [15,225]. In humans, the Moco deficiency causes the loss of SO and leads to early death in neonates. This occurs because the loss in SO activity results in the accumulation of sulfite (an intermediate product of the cysteine degradation), ultimately causing severe neurological damage among other disorders[217]. The assimilatory nitrate reductase catalyzes the reduction of nitrate to nitrite in autotrophic organisms (plants and fungi) and is completely different from the bacterial nitrate reductases[15].

In plants, as demonstrated for the SO of *Arabidopsis thaliana*, this enzyme removes the excess of sulfite produced during the sulfur assimilation. Once located in the peroxisomes, SO does not react with cytochrome c and uses, instead, oxygen as final electron acceptor[229]. In opposition, the *S. novella* SDH cannot transfer the electrons to molecular oxygen that are produced during the metabolization of the sulfite, formed during dissimilatory oxidation of reduced sulfur

compounds[228]. Enzymes of the SO family form dimeric structures: homodimers, in animals and plants, and heterodimers in the case of bacterial SDH. The cofactor composition is also variable, with the animal SO containing Moco and one b-type heme in each domain; the SO from plants containing Moco; and the SDH, containing Moco and a c-type heme in each subunit[15].

Recently, a new type of molybdoenzyme has been identified in eukaryotes, the mitochondrial amidoxime reducing component (mARC). mARC reduces N-hydroxylated substrates (amidoximes, N-hydroxy-sulfonamides and N-hydroxy-guanidines) into their active amino forms. Although the knowledge about these enzymes is still limited, they were classified as members of the SO family due to the presence of a conserved cysteine as a putative ligand of the Mo atom [224,230].

The **DMSOR family** is the more diverse in terms of structure and cofactor composition, having in common two equivalents of MGD (called bis-MDG) bound to Mo at the active site, which adopts a trigonal prismatic geometry. This variety is patent by the structures determined so far, ranging from simple enzymes as the *Rhodobacter* DMSOR, which is a monomer harboring only Moco, to very complex structures, as the *E. coli* FDH-N that, besides Moco, also contains several Fe/S and hemes[7]. The periplasmic DMSOR of *R. sphaeroides*[231] together with the *R. capsulatus*[232] were the first members of the family to be structurally characterized. In general, the members of the DMSOR family catalyze the oxygen transfer to/from a lone pair of the substrate, except for the formate dehydrogenases (FDH). The exceptional character of the FDH is related with the capability of catalyzing the oxidation and reduction of formate to carbon dioxide, but also its ability to incorporate either molybdenum or tungsten at the active site. Despite the differences, the overall fold of the DMSOR crystal structures available is very similar[15,233].

3.1.2.1. Xanthine oxidase family

The enzymes of the XO family are the best characterized mononuclear Mo-containing enzymes. With a few exceptions, they catalyze the hydroxylation of different types of substrates (such as aldehydes and aromatic heterocycles) according to the following reaction [15,234]:

$$\mathbf{RH + H_2O \rightarrow ROH + 2H^+ + 2e^-}$$

Members of this family are found in all types of life and include the well-known xanthine oxidoreductase (XOR) and aldehyde oxidase (AOX) from animals, and xanthine dehydrogenase, CO dehydrogenase and nicotinate dehydrogenase from prokaryotes. The bovine XO (*Bt*XO) is one of the most studied enzymes, purified since 1924 from cow's milk[224]. The XO is responsible for the conversion of hypoxanthine to xanthine, and xanthine to uric acid. Inherited XOR deficiency leads to xanthinuria, that can cause multiple organ failure due to the deposition of xanthine[235]. The XOR are extremely broadly distributed in nature, and only a few organisms oxidize xanthine

using alternative pathways. For example, the *Aspergillus nidulans* uses a Fe^{2+}/α-ketoglutarate hydroxylase to convert xanthine to uric acid. The mammalian aldehyde oxidase is also an important enzyme involved in the metabolism of several aldehyde compounds and in the biotransformation of drugs and xenobiotics [236,237].

All the molybdoenzymes that belong to this family contain redox-active centers in addition to the Moco, generally with one or two [2Fe-2S] clusters and one FAD. On the other hand, the aldehyde oxidoreductase from *D. gigas* (*Dg*AOR) lacks FAD and the 4-hydroxybenzoyl-CoA reductase from *Thauera aromatic* (*Ta*HBCR)[238] has an additional redox-active center, a [4Fe-4S] cluster. The most commonly used electron acceptors are O_2 or NAD^+, which react at the FAD site, while the substrate reacts at the molybdenum center [7,224]. The active site contains a $Mo^{VI}OS/O(OH)$ nucleus with no covalent bonding of the metal to the polypeptide chain, in contrast to the other families.

Members from this family are present either as heterodimers ($(\alpha\beta)_2$) (XDH from *R. capsulatus* – Figure 3.2) or as homodimers (α_2) as *Dg*AOR. In XDH from *Comamonas acidovorans*, for example, the two [2Fe–2S] clusters and FAD are found in one subunit, while the Mo-MPT cofactor is found in a second. In the XdhABC from *V. atypica*, a heterotrimer ($\alpha\beta\gamma$), the two [2Fe–2S] cofactors, the FAD cofactor and the MCD cofactor are each found in separate subunits. Contrarily to what has been stated for other molybdoenzymes, such XDH from *R. capsulatus*[239], for the heterotrimers *V. atypica* XdhABC, the dimerization is not a requisite for the Moco insertion and, consequently protein activation[7].

In 1995, Romão *et al* solved the first X-ray structure for XO family member, the homodimeric aldehyde oxidoreductase from *D. gigas*[240]. As previously mentioned, in *Dg*AOR the FAD is missing and replaced by a connecting segment. Beyond the MCD cofactor, *Dg*AOR contains two [2Fe-2S] in a single polypeptide chain. This protein is an exception in the XO family, presenting in its catalytically competent form an equatorial oxo ligand instead of the sulfido ligand[241]. Another enzyme that the crystal structure has been solved and presents an exceptional feature is the *Ta*HBCR[238]. This $(\alpha\beta\gamma)_2$ heterotrimeric protein catalyzes the irreversible removal of a phenolic hydroxy group from 4-hydroxybenzoyl-CoA, forming benzoyl-CoA and water. This enzyme is, so far, a unique case of the XO family, since it is a reductase rather than a dehydrogenase or oxidase, and contains a [4Fe–4S] cluster[7,224]. The cluster is found in a 41-amino acid residues insert in the FAD subunit and has been suggested to be involved in the transfer of electrons from ferredoxin for the reduction of 4-hydroxybenzoyl-CoA to the MCD active site[238,242].

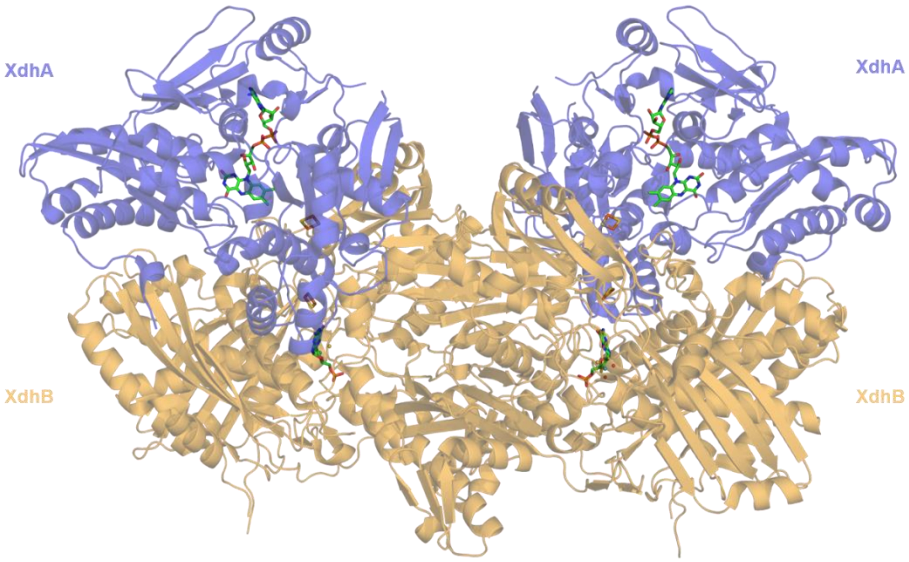


Figure 3.2. Cartoon representation of the xanthine dehydrogenase from *R. capsulatus*. The [2Fe-2S] and FAD cofactors of XDHA (blue) and the Mo-MPT of XDHB (orange) are shown as sticks.

3.1.2.1.1. Periplasmic aldehyde oxidoreductase and its chaperone

The periplasmic aldehyde oxidoreductase (PaoABC) from *E. coli* belongs to the XO family. Kinetic characterization of the enzyme showed that PaoABC converts a broad spectrum of aldehydes, with a preference for aromatic aldehydes, using molecular oxygen as the terminal electron acceptor[243]. Aromatic aldehydes are common in nature, present in plants and fruits (e.g., cinnamaldehyde), but also arise from fuel sources and air pollutants. At high doses, these molecules can act as antimicrobial agents and are often used as food preservatives. PaoABC degrades cinnamaldehyde quickly, as seen by the high catalytic rate determined experimentally (k_{cat} of 84 ± 5s^{-1}) [243]. In 2009, Neumann *et al* showed that a complete growth inhibition of *E. coli* devoid of genes derived from the *paoABCD* operon was achieved through the addition of cinnamaldehyde, probably damaging the bacterial cell surface[243]. This data suggests that PaoABC is part of the detoxification system of *E. coli* with an important protective role against aromatic aldehydes. This 135 kDa enzyme is composed by three subunits ($\alpha\beta\gamma$) with a large Moco-containing PaoC subunit (78.1 kDa), a medium FAD- containing PaoB subunit (33.9 kDa), and a small 2×[2Fe-2S]-containing PaoA subunit (21.0 kDa). Additionally, the PaoA contains a twin arginine protein transport (Tat) leader peptide for translocation to the periplasm of *E. coli*[243]. The amino acid sequences of the three subunits of PaoABC show significant similarities to enzymes of the xanthine oxidase (XO) family (30−40% identity). The structurally characterized members of this family with higher sequence homology with PaoABC are bovine milk XO (*Bt*XO)[237], XDH from *R. capsulatus*[244], CO dehydrogenase from *Oligotropha carboxidovorans*[245] and *Hydrogenophaga pseudoflava*[246], quinoline oxidoreductase from *Pseudomonas putida* (QoR) [247], and *Ta*HBCR (described previously)[238]. PaoABC is the first enzymes identified from *E. coli*

that binds the MCD form of Moco [243]. So far, almost all the characterized *E. coli* molybdoenzymes belong to the DMSOR family and have been shown to bind bis-MGD. The only enzyme that does not belong to this class is the *E. coli* MsrP protein (formerly called YedY), which belongs to the SO family and binds the Mo-MPT form of Moco[248]. In 2015, Gennaris *et al* showed the involvement of MsrP in repairing proteins containing methionine sulfoxide in the bacterial cell envelope [249].

The catalytic behavior of PaoABC is very different from the other molybdoenzymes of the XO family. For this protein, under steady state conditions, the turnover is maximal at pH 4 for the negatively charged ferricyanide and at pH 9 for a positively charged osmium complex [250]. The different pH-behavior depending on the electron acceptor and ionic strength conditions captured the attention of the biosensors community, revealing the urgency to clarify how the reactions occur. Recently, Badalyan *et al* showed the great potential of PaoABC for bioanalytical application and development of bioelectronic devices, namely for the detection of benzaldehyde and the neurotransmitter γ-aminobutyric acid (GABA) [251,252]. Benzaldehyde is commonly used in several areas, including the pharmaceutical and food industry. The scientific interests to detect this aromatic aldehyde is related to the experimental evidence of its carcinogenicity in mice and cytotoxicity in humans[251].

One of the most evident structural characteristics of the molybdoenzymes is the localization of the active site. Typically, the Moco is deeply buried within the protein structure, revealing the necessity of proteins that may act as chaperones to insert the cofactor and facilitate the proper folding of the target proteins after Moco insertion[253]. In XO family, the XdhC-like molecular chaperones are responsible for the maturation and protection of Moco (named after the well-characterized *R. capsulatus* XDH, XdhC). The XdhC binds Moco produced by MoeA/MogA and protects it prior to insertion of the terminal sulfur ligand by the NifS4[254–256]. After Moco sulfuration, XdhC dissociates from the NifS4 and inserts the cofactor into the XdhB subunits of the XDH heterotetramer $(\alpha\beta)_2$ [256]. The PaoD is the corresponding molecular chaperone of PaoABC and is encode by the *paoABCD* operon [243]. This independent 35 kDa protein was shown to be essential for the insertion of sulfurated MCD into PaoABC. Additionally, PaoD facilitates the sulfuration and insertion of an MCD cofactor rather than a Mo-MPT cofactor. Previous studies showed that PaoD interacts with CTP:molybdopterin cytidylyltransferase MocA and receives the MCD cofactor from it. The correct expression of the *paoABCD* operon is a requirement to obtain an active PaoABC enzyme[212,257].

3.2. Structural studies on PaoD

3.2.1. Experimental procedure

The *E. coli* PaoD was cloned and expressed by Viola Schwuchow (from Leimkühler's group). The reagents were purchased from Sigma-Aldrich unless stated otherwise.

3.2.1.1. Purification protocol

PaoD was expressed and purified using the procedure described by Neumann *et al* [212]. *E. coli* BL21(DE3) cells were transformed with plasmid pMN87. For expression, LB medium was inoculated with 1:100 overnight culture and incubated at 303 K until an OD at 600 nm of 0.3–0.5 was achieved. The expression was induced with 100 mM Isopropyl b-D-1- thiogalactopyranoside. After 5 h of growth, the cells were harvested and the cell pellet was resuspended in 50 mM phosphate buffer and 300 mM NaCl, at pH 8.0 (10 mL of buffer per liter of expression culture). Cell lysis was achieved after two passages through a TS Series Benchtop cell disruptor at 1350 bar in the presence of 1 mg/mL of DNase I. The cleared lysate was applied to a Ni-tris(carboxymethyl)ethylenediamine (Ni-TED – from Macherey-Nagel) column with 0.4 mL of resin per liter of culture. The column was washed with imidazole solutions at two different concentrations (10 and 20 mM) and PaoD was eluted with 250 mM imidazole in 50 mM phosphate buffer and 300 mM NaCl, pH 8.0. The buffer was exchanged to 50 mM phosphate buffer and 300 mM NaCl, pH 8.0 by size exclusion chromatography using a PD MultiTrap G-25 column (GE Healthcare) according to the manufacturer instructions. However, due to the high number of salt crystals obtained when phosphate buffer was used in crystallization, this buffer was abandoned in subsequent experiments and replaced by 50 mM Tris-HCl, 1 mM EDTA and 300 mM NaCl, pH 8.0. The purity of PaoD was determined by SDS/PAGE using Coomassie Brilliant Blue staining – Figure 3.3.

The purified protein showed a single band on Coomassie brilliant blue stained SDS polyacrylamide gel with a molecular mass of 36 kDa, which is in correspondence with the calculated mass of 34.8 kDa. Our collaborators show by size exclusion chromatography (Superdex 75), that PaoD is eluted with a calculated size of 71 kDa [258]. The observed elution position of native PaoD reveals that it exists in its native state as a dimer in solution. Only a small portion of the protein eluted as a tetramer from the size exclusion column, showing a small tendency to form larger aggregates.

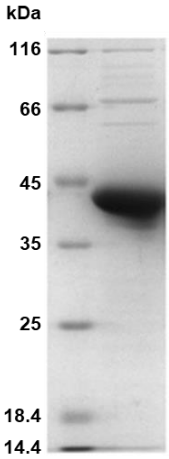


Figure 3.3. 12% SDS/PAGE of the purified PaoD after Ni-TED chromatography. Protein molecular weight marker Pierce™ Unstained Protein MW Marker (ThermoFisher Scientific). Running conditions: 20 mA during 1 hour at 298 K.

The concentration of the purified PaoD dimer was determined from the absorbance at 280 nm, using an extinction coefficient of 33920 $M^{-1}cm^{-1}$. This extinction coefficient was calculated using the bioinformatic tool ProtParam from the ExPASy portal [259].

3.2.1.2. Dynamic light scattering studies

DLS yields information about sample homogeneity and the size distribution of the hydrodynamic diameter of the protein complexes. To study the effect of different ionic liquids (IL), 0.4 M C_4mimCl and C_2OHmimPF_6, and additives (1 mM EDTA, 1 mM DTT, 1% Triton X-100 and 300 mM NaCl) in the protein stability, the protein buffer was changed to 50 mM phosphate buffer pH 8.0 or 50 mM Tris-HCl pH 8.0 by size exclusion chromatography using a PD MultiTrap G-25 column. PaoD at 0.7 mg/ml was incubated with different additives for 16 hours at 277 K and, in the case of ionic liquids, for 64 hours at 277 K. Before DLS measurements, all the solutions were centrifuged at 10000 rpm for 30 min and filtered through a microfilter with a pore size of 0.2 mm (Vivaspin® 500, Sartorius Stedim Biotech). The measurements were performed in an SZ-100 Nanopartica Series Instruments (Horiba Scientific, Kyoto, Japan) at 298 K in plastic disposable cells (four opening) and the detector positioned at 90°. The Z_{average} and autocorrelation curves presented were calculated from the average of 4 runs with 120 s each by SZ-100 software for windows.

3.2.1.3. Saturation transfer difference (STD) NMR

Nuclear magnetic resonance (NMR) spectroscopy is a powerful tool to study interactions of small ligands with biological macromolecules, such as proteins or nucleic acids. The saturation transfer difference NMR (STD-NMR) experiment has been used for some years to characterize protein-ligand complexes. The STD-NMR experiment is based on the nuclear Overhauser effect and in the observation of the ligand resonance signals. It can be used as a screening technique or as a

tool for identifying ligand moieties important for binding. The STD-NMR is based on the fact that for a weak-binding ligand (K_D ranging from 10^{-8} mol. L^{-1} to 10^{-3} mol. L^{-1}), there is exchange (equilibrium) between the bound and the free ligand state [260]. An STD experiment involves subtracting a spectrum in which the protein was selectively saturated (on-resonance spectrum) with signal intensities I_{SAT}, from one recorded without protein saturation (off-resonance spectrum), with signal intensities I_0. The STD effect was calculated by $\frac{I_0 - I_{STD}}{I_0}$, in which $(I_0 - I_{STD})$ is the peak intensity in the STD spectrum and I_0 is the peak intensity in the off-resonance spectrum. The STD intensity of the largest STD effect was set to 100% as a reference and the relative intensities were determined [260–262].

The STD-NMR experiments were performed with PaoD in 50 mM Tris-HCl, 1 mM EDTA and 300 mM NaCl, pH 8.0 and two different ionic liquids, $[C_4mim]Cl$ and $[C_2OHmim]PF_6$. The final concentrations of protein and ionic liquids were *ca* 30 mM and 3 mM, respectively.

All STD-NMR experiments were performed at 310 K on a Bruker Avance III spectrometer operating at 600 MHz, with a 5 mm triple resonance cryogenic probe head. The STD-NMR spectra were acquired with 1024 transients in a matrix with 32 k data points in $t2$ in a spectral window of 12019.23 Hz centered at 2814.60 Hz. Excitation sculpting with gradients was employed to suppress the water proton signals. A spin lock filter ($T1\rho$) with a 2 kHz field and a length of 20 ms was applied to suppress protein background. Selective saturation of protein resonances (on-resonance spectrum) was performed by irradiating at 2300 Hz using a series of 40 Eburp2.1000 shaped 90u pulses (50 ms, 1 ms delay between pulses), for a total saturation time of 2.0 s. For the reference spectrum (off-resonance) the samples were irradiated at 20000 Hz. Proper control experiments were performed with the reference samples to optimize the frequency for protein saturation (20.5 ppm) and off-resonance irradiation, to assure that the ligand signals were not affected.

3.2.1.4. Crystallization and data collection

Crystallization trials of PaoD were prepared with the protein in 50 mM Tris-HCl, 1 mM EDTA and 300 mM NaCl, pH 8.0, or 50 mM phosphate buffer and 300 mM NaCl, pH 8.0. The protein samples were thawed in the presence of 0.2 or 0.4 M of the ionic liquid (IL) $[C_4mim]Cl$ as well as of IL $[C_2OHmim]PF_6$ (Solchemar). When using either IL at 0.4 M, no precipitation was observed and the protein could be concentrated by centrifugation using a Vivaspin 2 ultrafiltration device (Sartorius Stedim Biotech S.A.). Protein solutions of 5 mg/mL could thus be obtained when using both IL at 0.4 M and were used for crystallization assays employing the vapor diffusion method. Several commercial screenings were tested, namely JBScreen Classic 1–10 (Jena Bioscience), MemStart (Molecular Dimensions) and 80 conditions in-house screen (based on the screen of Jancarik *et al*[182]). Crystallization drops of 0.2+0.2 µL were set- up using the automatic protein crystallization system Oryx8 (Douglas Instrument).

From the several crystallization conditions tested only two generated protein diffracting crystals when the Tris-HCl buffer pH 8.0 was used: in condition 1 protein thawing was done in the presence of 0.4 M [C_4mim]Cl and ammonium sulfate at 2.2 M was used as precipitating agent; condition 2 contained 12% PEG 4K, 0.1 M Tris-HCl pH 8.5 and 10 mM cysteine and the protein sample had been thawed in the presence of [C_2OHmim]PF_6 (Figure 3.4). The reducing agent L-cysteine has proven to be a good choice to improve condition 2 since, in its absence, poorly diffracting crystals were obtained. Other crystallization additives (Additive Screen, Hampton Research) were tried, without success. Both crystallization conditions generated small crystals with the same morphology within one and two months (Figure 3.4), respectively, but were very difficult to reproduce.

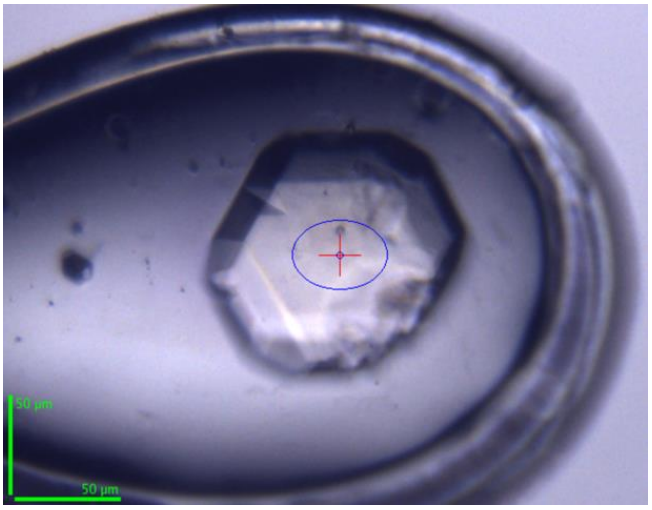


Figure 3.4. PaoD crystal. The protein was thawed in presence of 0.4 M of [C_2OHmim]PF_6. The crystallization condition was 12% PEG 4 K, 0.1 M Tris-HCl pH 8.5 and 10 mM cysteine. The same crystal morphology was obtained for the $P6_122$ crystals grown from ammonium sulfate. The image was captured on the synchrotron beamline ID23-1 (ESRF, France).

Scale-up attempts were performed in 24 well crystallization plates (Molecular Dimensions) using 1+1, 1+2 or 2+1 µL drops of protein+precipitant, with no success. Nevertheless, synchrotron data collection could be achieved using paratone and glycerol as cryoprotectants, before stabilization with a harvesting buffer. Analysis of the diffraction patterns suggested that paratone is not an adequate cryoprotectant and very poor diffraction was observed. Comparatively, crystals that were flash-frozen after a quick soak with a crystallization solution supplemented with 30% (v/v) glycerol showed a better diffraction pattern and a complete dataset could be collected. Due to the scarcity of crystals, no other cryo solutions or measurements at room temperature were performed. The crystals obtained with condition 1, where ammonium sulfate was used as precipitant and the protein thawed in the presence of 0.4 M of [C_4mim]Cl, diffracted to a maximum resolution of 3.39 Å (Figure 3.5.A). For the crystallization condition 2 using PEG as the precipitant agent and PaoD thawed in the presence of 0.4 M [C_2OHmim]PF_6, crystals diffracted to a maximum resolution of 2.29 Å (Figure 3.5.B). These datasets were analyzed and processed with the programs Mosflm[263], Pointless[264] and Scala[265] that showed that crystals obtained with

condition 1 belong to $P6_122$ space group while the ones from condition 2 belong to $P3_121$ space group.

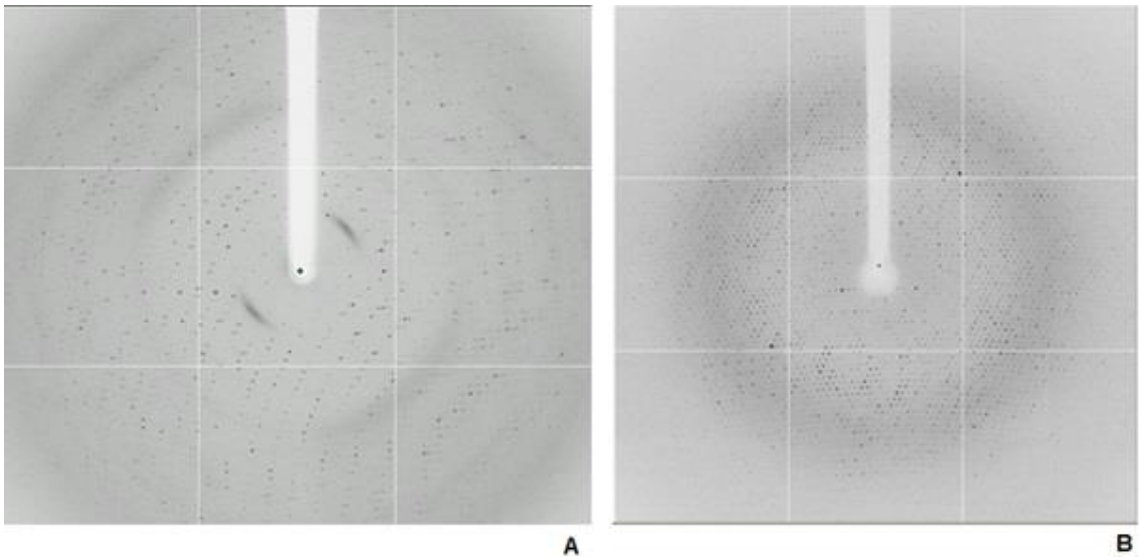


Figure 3.5. Diffraction pattern of two different PaoD crystal forms. A. Crystal obtained with condition 1 where the protein was thawed in the presence of 0.4 M $[C_4mim]Cl$ and ammonium sulfate was used as crystallization agent. B. Crystal obtained with condition 2 where the protein was thawed in the presence of 0.4 M $[C_2OHmim]PF_6$ and using PEG 4 K as crystallization agent.

3.2.1.5. Preliminary crystallization (and structural NMR) studies of other related proteins

The results presented in this Chapter, sections 3.2.2 and 3.3.2, allowed the determination of the X-ray structures of various proteins involved in the molybdenum cofactor biosynthesis. Several crystallization trials were performed for the proteins presented in Table 3.2.

A lot of effort was applied to accomplish this task, and PaoD, NifS4, and MocA have been crystallized. The first problems encountered were the lack of solubility of PaoD that could be improved by the use of ionic liquids[258,266] (section 3.2.2.1).

As to the crystals of NifS4 and MocA obtained they were twinned or diffracted very poorly. The crystals are under optimization.

The FdsD is a 9 kDa protein that, together with FdsC, is essential to produce an active FDH. The specific role of this small protein remains unclear, and the structural studies are essential for elucidation for the mechanism determination. Since the FdsD has a low molecular weight and there is no homology model that allows structure determination by molecular replacement, labeled proteins ^{15}N-FdsD and $^{13}C^{15}N$-FdsD were produced to perform NMR experiments. All spectra were collected and are under analysis. The first 1D and 2D NMR spectra reveal that FdsD has a proper folding and the number of peaks in the ^{1}H-^{15}N HSQC is coherent with the number of

expected residues (94 amino acids) – Figure 3.6. Structure calculations are being performed by Dr. Jorge Dias and Prof. Eurico Cabrita (UCIBIO@Requimte, FCT-UNL).

Table 3.2. Proteins involved in the molybdenum cofactor biosynthesis and maturation subjected to crystallization assays.

Protein name and organism	Function	Maximum resolution of the obtained crystals (Å)
MocA[223] *Escherichia coli*	CMP transfer to Mo–MPT, formation of MCD	7.0
MobA[256] *Escherichia coli*	Formation of bis-Mo–MPT, GMP transfer to bis-Mo–MPT, formation of bis-MGD	NA
PaoD[243,258] *Escherichia coli*	Stabilization of the MCD (S) and insertion into their respective target proteins (XDH or PaoABC).	2.29
XdhC[253] *Rhodobacter capsulatus*		NA
NifS4[255] *Rhodobacter capsulatus*	Sulfur transfer to Moco, formation of MCD(S)	3.0
FdsC[267] *Rhodobacter capsulatus*	Protect the bis-MGD cofactor from oxidation before its insertion into FdsA.	NA
FdsD[267] *Rhodobacter capsulatus*	Unspecific function	NA

NA – not available

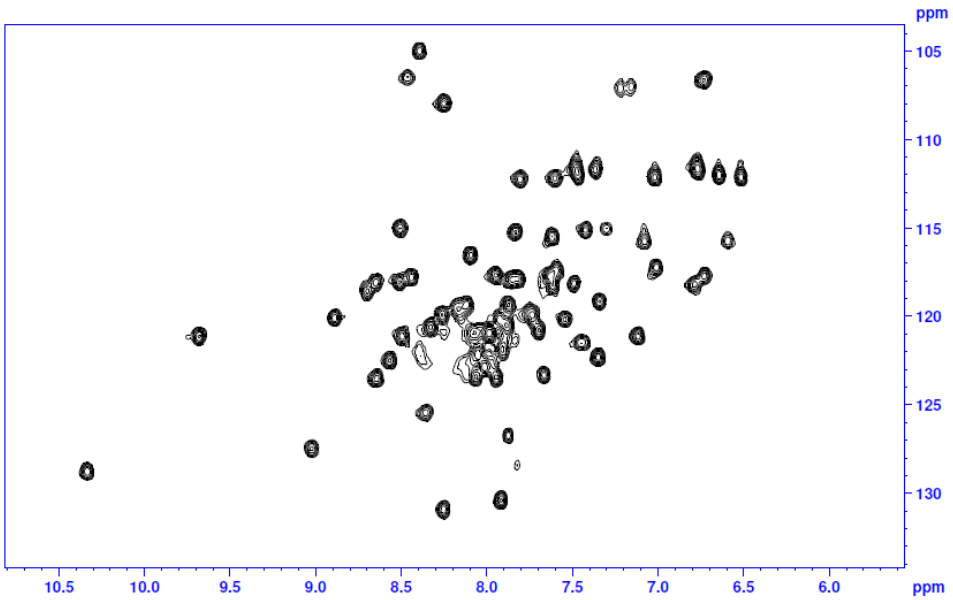


Figure 3.6. ^{1}H-^{15}N HSQC spectrum of the FdsD. 200 µM ^{15}N-labeled FdsD samples spectrum acquired on a 600 MHZ NMR spectrometer with cryoprobe, at 293 K.

The following results and discussion are dedicated to the chaperone PaoD.

3.2.2. Results and discussion

3.2.2.1. Effect of ionic liquids on protein stability

Low stability is a detrimental characteristic for any structural study of biomacromolecules. PaoD is an unstable, medium size protein that tends to aggregate, especially after a cycle of freezing/thawing. Low expression levels together with the difficulty in achieving high protein concentrations compromised crystallization experiments. To overcome this, several compounds commonly known as protein stabilizers, have been used, such as DTT, Triton X-100 and EDTA. In recent years, there has been an increasing attention to the use of ionic liquids (ILs) to increase protein stability[268,269] as well as crystallization additives[266,270,271]. ILs are a class of organic salts that melt below ~100 °C and have an appreciable liquid range. An IL is entirely composed of highly asymmetric positive and negative ions and the combinations can be customized for specific applications [272,273].

The tendency of protein samples to aggregate is a major drawback for crystallization since heterogeneity may hamper good crystal packing and therefore the formation of well-ordered crystals. DLS measurements were performed in order to study the effect of different additives (EDTA, DTT, Triton X-100 and NaCl) and ionic liquids ([C_4mim]Cl and [C_2OHmim]PF_6) upon protein stability. Table 3.3 summarizes the $Z_{average}$, obtained from the autocorrelation functions present in Figure 3.7 and 3.8, and the polydispersity index (PI) obtained for each additive and buffer. The $Z_{average}$ obtained for PaoD in the two buffers tested, 50 mM Tris-HCl (pH 8.0) and 50 mM Phosphate (pH 8.0) are high (over 100 nm) indicating that medium-large size aggregates are formed under these conditions. Common additives such as EDTA, DTT or NaCl exhibit either no effect (as for DTT) or drastically decrease protein stability, which is denoted by the very large $Z_{average}$ values observed (500 to 5000 nm). In the presence of Triton X-100 smaller aggregates of PaoD are formed in Tris-HCl buffer, but not in phosphate buffer. The data show that the best additives tested are the two IL since both gave smaller $Z_{average}$ values in the two buffers. The best results were obtained for [C_2OHmim]PF_6 in Tris-HCl buffer where $Z_{average}$ values are of 48 nm, even though this corresponds to small size aggregates for a 35 kDa protein as PaoD. The effect of the two IL has been tested over time and, in Tris-HCL buffer, even after 64 hours of incubation, the $Z_{average}$ values observed are within the same order of magnitude. For PaoD, all the tested additives in phosphate buffer (50 mM, pH 8.0), the $Z_{average}$ increases in comparison with the results for 50 mM Tris-HCl, pH 8.0. This indicates that the latter is better suited for this protein. Even in the case of phosphate buffer, the addition of the ILs leads to a reduction of the particle size of *ca* five times when compared with the protein in buffer alone, showing the stabilization properties of the two ILs.

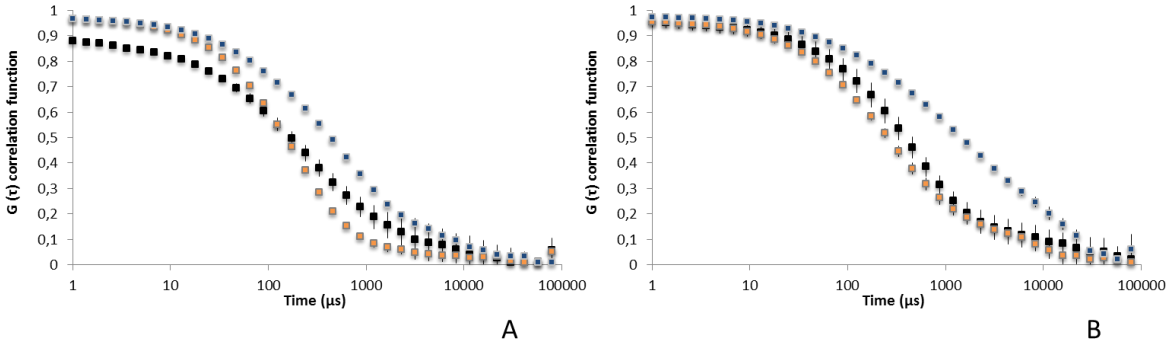


Figure 3.7. Autocorrelation curves for PaoD in presence of different ionic liquids after 16 hours of incubation. (A) Protein in 50 mM Tris-HCl, 300 mM NaCl and 1 mM EDTA pH 8.0 (blue), with 0.4 M [C_4mim]Cl (black) and 0.4 M [C_2OHmim]PF_6 (orange). (B) Protein in 50 mM Phosphate buffer and 300 mM NaCl pH 8.0 (blue), with 0.4 M [C_4mim]Cl (black) and 0.4 M [C_2OHmim]PF_6 (orange).

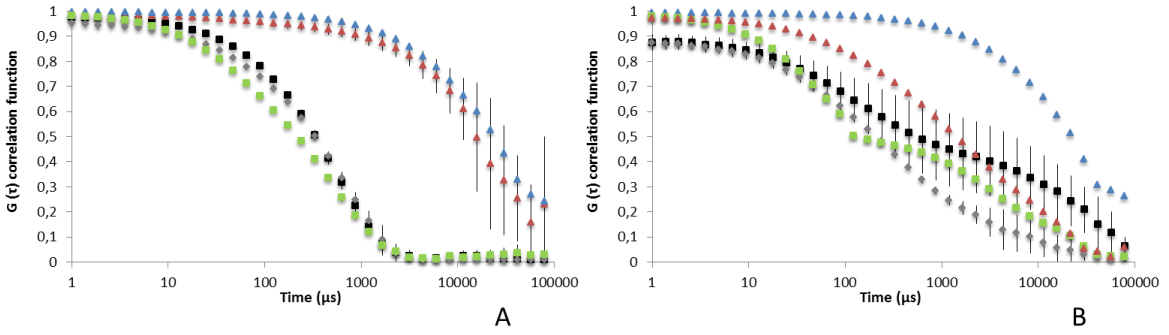


Figure 3.8. Autocorrelation curves for PaoD in presence of different additives after 16 hours of incubation. (A) Protein in 50 mM Tris-HCl pH 8.0 (black), with 1 mM DTT (grew), 300 mM NaCl (pink), 1 mM EDTA (blue) and 1% Triton X-100 (green). (B) Protein in 50 mM Phosphate buffer pH 8.0 (black) with 1 mM DTT (grew), 300 mM NaCl (pink), 1 mM EDTA (blue) and 1% Triton X-100 (green).

Table 3.3. Comparison between $Z_{average}$ and polydispersity index for PaoD with different additives and for the two IL, after 16 and 64* hours of incubation. Data in parenthesis correspond to the polydispersity index.

	$Z_{average}$ **and polydispersity index (PI)**	
Buffer **Additive**	**Tris-HCl (50 mM, pH 8.0)**	**Phosphate buffer (50 mM, pH 8.0)**
0.4 M [C_4mim]Cl[§]	105.7±54.8 nm (0.43±0.10) 103.3±13.4 nm (0.50±0.07)*	165.1±38.6 nm (0.38±0.06) 300.3±46.6 nm (0.43±0.04)*
0.4 M [C_2OHmim]PF_6[§]	48.8±1.1 nm (0.35±0.03) 89.0±5.5 nm (0.18±0.03)*	116.3±25.3 nm (0.37±0.03) 576.6±4.5 nm (0.57±0.11)*
1 mM DTT	99.8±10.5 nm (0.58±0.14)	248.7±52.8 nm (0.34±0.02)
1 mM EDTA	5391.3±1625 nm (1.11±0.30)	4234.9±432.8 nm (1.76±0.40)
1% Triton X-100	62.3±4.9 nm (0.56±0.05)	615.5±136.2 nm (0.54±0.07)
300 mM NaCl	2844.2±470 nm (1.81±0.2)	557.0±3.0 mn (0.50±0.02)
300 mM NaCl and 1 mM EDTA	175.8±39.1 nm (0.40 ±0.05)	-
-	101.7±8.2 nm (0.47±0.05)	630.4±60.5 nm (0.53±0.19)

[§]Assays performed with buffer supplemented with 300 mM NaCl. The Tris-HCl pH 8.0 buffer was supplemented with 1 mM of EDTA.

In the absence of ionic liquids, the formation of large aggregates is obvious since protein precipitation is clearly visualized after 24 hours at 277 K. The fact that the presence of these stabilizing molecules seems to reduce the number of protein aggregates in solution could be the reason for the success in crystallization under these conditions.

3.2.2.2. Interaction of ionic liquids with PaoD and STD-NMR data

As described in the literature[270], either the anionic or cationic counterpart of the ILs can stabilize proteins through protein-ion interactions. To understand the putative interactions between the PaoD and the IL, responsible for the observed increase in stabilization (by DLS), STD-NMR experiments were performed.

STD-NMR is a robust method that can provide information about intermolecular interactions from the viewpoint of the small molecule, allowing the characterization of low-affinity interactions between small molecules and (bio)macromolecules[260,274]. The STD-NMR experiment is based on a transfer of saturation from the protein to the ligand, which is in a much higher concentration compared to the protein. By measuring this saturation transfer it is possible to identify the existence of a protein-ligand interaction and determine which part of the ligand is responsible for this interaction. STD-NMR experiments were performed with [C_2OHmim]PF_6 and [C_4mim]Cl, as

described in the experimental section, and the resulting STD-NMR spectra are presented in Figure 3.9.

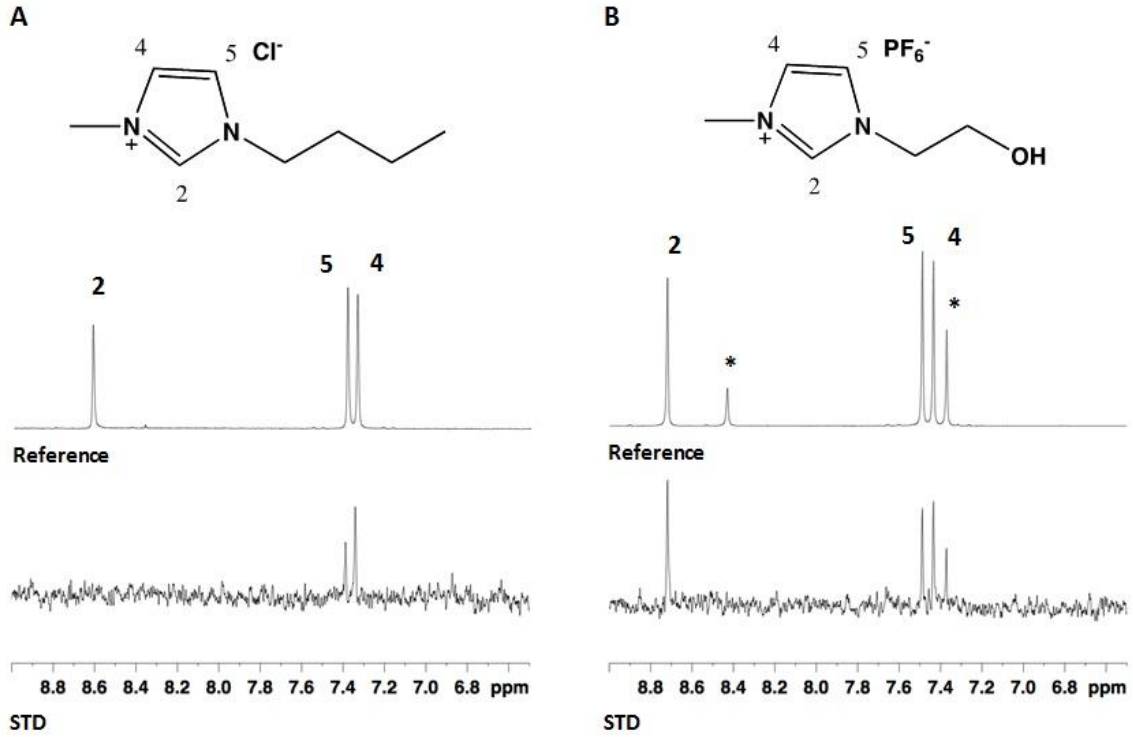


Figure 3.9. Expansion of the aromatic region of (A) the reference and the STD-NMR spectrum obtained with [C_4mim]Cl and (B) the reference and the STD-NMR spectrum obtained with [C_2OHmim]PF_6. *Marks the peaks of imidazole present in the protein sample.

As can be seen in Figure 3.9, STD responses from the aromatic protons of the cation were detected both for [C_2OHmim]PF_6 and [C_4mim]Cl. This result is a clear indication that there is an interaction between the protein and the ILs in solution. Since the pI of the protein, determined by sequence analysis using the ExPASy portal[259], is around 6.5 and at the working pH (8.5) the protein is negatively charged, we anticipate that electrostatic forces can be the main driving force for the interaction. More interesting is the fact that no STD signals from the methyl group or alkyl protons of the imidazolium ring substituents could be detected. The absence of signals from these moieties and the comparison of the intensities between the reference and the STD spectra seem to suggest that the interaction between the IL cation and the protein has some degree of specificity and directionality. This directionality should be similar to the one found between the IL cation and the anion, which occurs preferentially with the more acidic ring protons 2, 4 and 5, that are able to participate in hydrogen bonds.

The major difference between the two ILs is the absence of a response from the more acidic proton 2 in the case of [C_4mim]Cl and an overall higher relative STD intensity for [C_2OHmim]. At this point, no hypothesis explains this result completely. However, the charge dispersion and polarity of the cation may be responsible for the differences. For weak binding ligands, the intensity of the STD can provide qualitative information regarding the relative affinity and

specificity of the different ionic liquids, if the experiments are performed under the exact same conditions of concentration and IL:protein ratio. Under these conditions, a higher intensity of the STD response can be due to a higher affinity, a less specific interaction or both since these factors will increase the efficiency of the saturation transfer process. Since [C_2OHmim] is smaller and more polar than [C_4mim] this can favor a stronger and less specific interaction with the protein that would explain the higher intensity of the STD response for [C_2OHmim] when compared to [C_4mim]. Therefore [C_2OHmim] can present a more extended interaction surface than [C_4mim], which, most likely, is contributing to an increase in PaoD stability, thus explaining why protein precipitation is avoided upon thawing in the presence of this IL. The anion may also play a role in the interaction since for the two ILs, different degrees of solvation of the ion pairs are expected by the buffer. This can have consequences in the availability of the cation to participate in interactions with the protein but a more extended investigation with different combinations of cations/anions is required to clarify the relative contribution of the anion [32].

Since the STD-NMR experiment is a ligand observed experiment, the results obtained do not allow to identify the moieties of the protein responsible for the interaction. However, the pattern of the interaction observed suggests that charged negative side chain residues, prone to participate in hydrogen bonds, could be responsible for the directionality of the interaction observed, explaining the lack of STD response from the other IL moieties.

3.2.2.3. Crystallographic data

The crystallization trials of PaoD were performed with a protein solution concentrated to approximately 5 mg/mL (in the storage buffer 50 mM Tris-HCl, 300 mM NaCl, 1 mM EDTA or 50 mM phosphate buffer and 300 mM NaCl, pH 8.0) in the presence of ionic liquids. Since protein precipitation occurs upon thawing of the sample (that at the time was send by our collaborators in dry ice), we decided to add the IL during this process, to a final concentration of either 0.2 or 0.4 M. In this way, the protein would thaw in the presence of the IL, which would help to prevent the precipitation and the procedure turned out to be very successful for [C_4mim]Cl and [C_2OHmim]PF_6 at 0.4 M. However, due to the high number of salt crystals obtained when phosphate buffer was used for crystallization, this solution was abandoned in subsequent experiments. After performing multiple commercial and in-house screenings, it was possible to obtain single, well diffracting crystals from two different crystal forms using two different crystallization agents: condition 1 with 2.2 M of ammonium sulphate and the protein thawed in the presence of [C_4mim]Cl and condition 2 with 12% PEG 4K, 0.1M Tris-HCl pH 8.5 and 10 mM cysteine, and the protein thawed with [C_2OHmim]PF_6 – Table 3.4.

Structure determination has not yet been accomplished due to the lack of a proper homology model. The protein was also labeled with selenomethionine for determining the phases by single- or multi-wavelength anomalous dispersion but, so far, no crystals were obtained. The crystals are very difficult to reproduce and the protein concentration is limited to 5 mg/mL. As future work, it

may be interesting to cleave the his-tags (currently the protein contains N- and C-terminal tags) and test other expression conditions and vectors.

Table 3.4. Data collection statistics for PaoD crystals. Data in parenthesis correspond to the highest resolution shell.

Crystal form	1	2
Crystallization condition	2.2 M Ammonium sulfate	12% PEG 4K, 0.1M Tris-HCl pH 8.5 and 10 mM cysteine
Ionic liquid	0.4 M [C_4mim]Cl	0.4 M [C_2OHmim]PF_6
	Data collection	
X-ray source	ID23-1, ESRF, France	
Wavelength (nm)	0.975	0.979
	Processing statistics	
Space group	$P6_122$	$P3_121$
Unit-cell parameters (Å)	$a = b$ =144.44, c =240.48 $\alpha = \beta = 90°, \gamma = 120°$	$a = b$ =106.41, c =237.41 $\alpha = \beta = 90°, \gamma = 120°$
Matthews parameter (Å^3/Dalton)	2.59 (4 molecules/ AU)	2.77 (4 molecules/ AU)
No. observed reflections	285511 (36291)	356767 (17503)
No. unique reflections	21154 (2917)	70445 (4395)
Resolution limits (Å)	86.59 – 3.39 (5.57 – 3.39)	60.04 – 2.29 (2.35 – 2.29)
Completeness (%)	99.0 (96.8)	99.8 (97.4)
R_{pim} (%)*	4.0 (15.7)	4.1 (10.7)
$< I/\sigma(I) >$	14.5 (5.6)	9.7 (3.2)
Multiplicity	13.5 (12.4)	5.1 (4.0)
Average mosaicity	0.77	0.45

[a] $R_{pim} = \sum_{hkl}[\frac{1}{N-1}]^{1/2} \sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hlk} \sum_i I_i(hkl)$, where N is the multiplicity measured.

3.3. Structural elucidation of the *E. coli* Periplasmic Aldehyde Oxidoreductase PaoABC

3.3.1. Experimental procedure

DNA cloning, expression, and purification of *E. coli* PaoABC were performed by Meina Neumann (from Leimkühler's group). Viola Schwuchow was also responsible for the site-directed mutagenesis, purification, and quantification of the PaoABC mutants (PaoC-R440H and PaoC-R440K).

3.3.1.1. Crystallization and data collection

The PaoABC from *E. coli* was expressed and purified according to the protocol described by Neumann *et al*[243]. The enzyme was concentrated by ultrafiltration (Vivaspin® 20, Sartorius) up to 20 mg/mL in 50 mM Tris-HCl pH 7.5, 250 mM NaCl, 1 mM EDTA. The final concentration was determined from the absorbance at 445 nm, using an extinction coefficient of 23686 $M^{-1}.cm^{-1}$ for the native enzyme[243]. The first crystallization screening experiments were performed at 293 K by hanging-drop vapor diffusion method with 1 µL of protein + 1 µL of the precipitant solution on a 24-well XRL plate (Molecular Dimension). Several commercial screenings were tested, namely PEG/Ion HT (Hampton Research), JBScreen Classic 1-10 (Jena Bioscience), and an in-house sparse matrix screen (based on the screen of Jancarik *et al*[182]– formulation in the Table A1, Appendix).

PaoABC crystallized only in one condition of the commercial screen PEG/Ion HT that contains 0.2 M ammonium iodide and 20% (w/v) polyethylene glycol (PEG) 3350 (in 1+1 µL drop). Yellow-brownish, plate shape crystals appeared within two days. However, the first datasets reveal that the crystals measured were twinned, with a twinning fraction[275] of 50%. To overcome this, several concentrations of ammonium iodide (between 0.1 M and 0.25 M), percentage of PEG 3350 (between 10% and 30%), proportions of drop and additives (Additive Screen, Hampton Research) were tested but without success. When the crystallization temperature was changed from 293 K to 277 K, the crystals took four days to appear reaching maximum dimensions of 0.08 × 0.2 × 0.2 mm in the same crystallization condition – Figure 3.10. The crystals were transferred into a new drop with a harvesting solution containing a higher percentage of PEG 3350 (22% (w/v). After 3-5 min incubation, the crystal was then flash-cooled in liquid nitrogen, using 30% (v/v) glycerol as a cryoprotectant, and maintained at 100 K under a stream of liquid nitrogen during data collection. Complete datasets were collected at PXIII (X06DA) beamline of the Swiss Light Source (SLS,

Switzerland) and Proxima I beamline at Soleil (France). The crystals diffracted up to 1.7 Å and 1.8 Å at Proxima I and PXIII beamlines, respectively. Both crystals belong to the centred monoclinic space group C2 and with a solvent content of *ca* 48%. Due to the highest resolution, the structure was solved using data collected at Soleil. The data was processed with iMOSFLM v.1.0.7[276] and SCALA[277] from CCP4 program package v. 6.3.0[185]. The data collection and processing statistics are presented in Table 3.5.

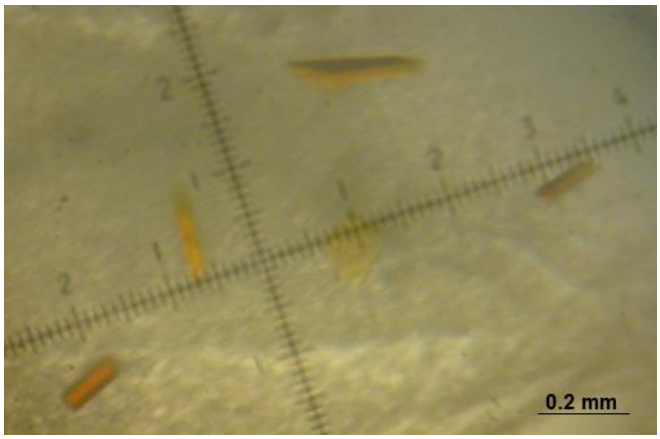


Figure 3.10. PaoABC crystals obtained in 0.2 M ammonium iodide and 20% (w/v) PEG 3350.

The same experimental procedure was performed to crystallized the two PaoABC mutants: PaoC-R440H and PaoC-R440K. Crystals of the PaoC-R440H mutant were obtained in the same condition of the wild-type after three days at 277 K. The crystals diffracted up to 2.3 Å and a complete dataset was collected in the BM14 beamline at ESRF (France). The data was processed with iMOSFLM v.1.0.7[276] and SCALA[277] from CCP4 program package v. 6.3.0[185]. The data-collection and processing statistics are presented in Table 3.5.

Crystals of the PaoC-R440K were also obtained in the same crystallization condition, although the majority were anisotropic or with poor-diffraction. The crystals are under optimization but so far without success.

Table 3.5. Crystallographic data of PaoABC wild-type and PaoC-R440H mutant from *E. coli.* Statistics for the highest-resolution shell are shown in parentheses.

	PaoABC wild-type	PaoABC/PaoC-R440H
Data collection parameters		
X-ray Source	Proxima I (Soleil, France)	BM14 (ESRF, France)
Detector	PILATUS 6M	PILATUS 2M
Wavelength (Å)	0.976	0.979
Processing statistics		
$\boldsymbol{a}$ **(Å)**	109.68	109.84
$\boldsymbol{b}$ **(Å)**	78.34	78.26
$\boldsymbol{c}$ **(Å)**	151.91	151.73
$\boldsymbol{\beta}$ **(˚)**	99.69	99.93
Space group	C2	C2
Molecules per AU	1	1
Matthews coefficient (Å³/Da)	2.39	2.39
Mosaicity (˚)	0.14	0.35
Resolution range (Å)	48.32 - 1.70 (1.73 - 1.70)	48.27 – 2.37(2.30 – 2.30)
$< \boldsymbol{I}/\boldsymbol{\sigma I} >$	9.69 (1.90)	8.7 (1.81)
$\boldsymbol{R_{pim}}$ **(%)**[a]	8.6 (69.5)	12.8 (65.1)
Multiplicity	4.4 (4.4)	7.2 (4.9)
Number of observed reflections	605876 (30312)	397602 (19057)
Number of unique reflections	137343 (6876)	55360 (3851)
Completeness (%)	98.53 (99.04)	98.0 (84.2)

[a] $\boldsymbol{R_{pim}} = \sum_{hkl}[\frac{1}{N-1}]^{1/2} \sum_i |\boldsymbol{I_i(hkl)} - \langle \boldsymbol{I(hkl)} \rangle| / \sum_{hlk} \sum_i \boldsymbol{I_i(hkl)}$, where $\boldsymbol{N}$ is the multiplicity measured.

3.3.1.2. Structure determination and refinement

Structure determination of wild-type PaoABC was performed with PHASER[186] using several molecular models according to sequence alignment homologies. The best models were chosen by comparing sequence identities of the individual subunits against the PDB database using BLAST[278]. The 4-hydroxybenzoyl-CoA reductase from *Thauera aromatica* (*Ta*HBCR, PDB code 1rm6)[238], Quinoline 2-Oxidoreductase from *Pseudomonas Putida* 86 (*Pp*QoR, PDB code 1t3q)[247] and Carbon monoxide dehydrogenase from *Hydrogenophaga pseudoflava* (*Hp*CODH, PDB code 1ffu)[246] were chosen to determine the solution for subunit A; *Ta*HBCR, hpCODH (PDB code 1ffu and 1ffv) for subunit B; and *Ta*HBCR, Desulfo-Xanthine Oxidase with xanthine (PDB code 3eub)[279] and reduced Xanthine Oxidase in complex with arsenite (PDB code 3sr6)[280] from *Bos taurus* for subunit C. A molecular replacement solution was obtained for the three subunits and clear electron density was observed for the entire protein as well as for the expected cofactors MCD, two distinct [2Fe-2S] clusters and one FAD, that had been removed from the search models. A density modification protocol was applied using DM[281] giving initial phases with *ca.* 0.7 mean figure of merit. Restrained refinement was performed with REFMAC 5.2[187] and inspection

of the electron density maps was carried out using COOT[188]. The last three programs are available at CCP4[185] suite of programs.

During refinement, extra electron density could be observed not far from the FAD site. Inspection of the $F_{obs} - F_{cal}$ and the anomalous maps (3.0σ and 5.0σ, respectively) revealed strong electron density peaks. The peaks were interpreted as corresponding to four iron atoms of a [4Fe-4S] center although its presence had not been anticipated for PaoABC. In the final stages of refinement, the R_{factor} and R_{free} converged to 13.7% and 16.6% for the wild-type enzyme.

The wild-type was used as search model to determine the PaoABC structure with a mutation in the R440H of the subunit C. After refinement, final values of 16.8% and 21.6% were obtained for R_{factor} and R_{free} factors, respectively. Geometrical validation and model improvement was carried out using PDBREDO[189] and several validation programs such as PROCHECK[282] and MOLPROBITY[283]. Analysis of both models (wild-type and mutant) showed that 98.0% of the protein residues are in the most favored or additionally allowed regions of the Ramachandran plot, while only 0.16% are in disallowed regions. Refinement statistics are summarized in Table 3.6.

Table 3.6. Structure refinement statistics for PaoABC wild-type and mutant PaoC-R440H

Refinement statistics		
	PaoABC wild-type	**PaoABC/PaoC-R440H**
R_{factor} **(%)**[a]	13.7	16.8
R_{free} **(%)**[b]	16.6	21.6
Average B-factor for all atoms (Å²)	21.20	25.30
RMSD from ideal geometry		
Bond lengths (Å)	0.015	0.010
Bond angles (°)	1.653	1.399

[a] $R_{work} = \sum ||F_{calc} - F_{obs}|| \,/\, \sum |F_{obs}| \times 100$, where F_{calc} and F_{obs} are the calculated and observed structure factor amplitudes, respectively.
[b] R_{free} is calculated for a randomly chosen 5% of the reflections.

Coordinates and observed structure factor amplitudes have been deposited in PDB under the accession codes **5G5G** for wild-type PaoABC and **5G5H** for PaoC-R440H mutant.

3.3.1.3. Small-Angle X-ray Scattering

SAXS data were collected at the EMBL beamline X33 at DESY in Hamburg[284]. The measurements were performed at 293 K and different concentrations ranging from 0.3 to 30 mg/mL were used. Data were recorded using a Pilatus 1 M pixel detector at a sample-detector distance of 2.7 m and a wavelength of 1.5 Å, covering the range of momentum transfer $0.01 < s < 0.6$ Å^{-1}. Sample solutions were circulated in a thermostated cuvette, positioned within a

vacuum chamber. Eight frames of 15 s each were collected, normalized to the transmitted intensity, and subsequently averaged using AUTOSUB[81]. The data were processed with the ATSAS package[85] using standard procedures, corrected for buffer contribution, and extrapolated to infinite dilution using the program PRIMUS[192]. The $I(0)$ and the R_g were evaluated using the Guinier approximation, but also computed from the entire scattering pattern using GNOM[88]. GNOM also provides the D_{max} and the $P(r)$. The program CRYSOL was used to compute the scattering from the known high-resolution models of *Ta*HBCR (PDB code 1rm6)[94]. The *ab initio* modeling programs DAMMIN[73] and DAMMIF[91] were employed for low-resolution shape generation, and 20 models were calculated in the slow mode, using standard settings. The program DAMAVER[75] was utilized to superimpose individual structures, and to determine the averaged and the most probable reconstruction. The *ab initio* model was superimposed with the high-resolution structure of 4-hydroxybenzoyl-Coa reductase using SUPCOMB[205].

3.3.2. Results and discussion

3.3.2.1. Overall structure

X-ray Crystallography Results

The crystallographic structure of periplasmic aldehyde oxidoreductase (PaoABC) was determined at a high resolution (1.7 Å) using synchrotron radiation data. Crystallographic results support that PaoABC is a heterotrimeric ($\alpha\beta\gamma$) enzyme embodying a set of redox centers involved in electron transfer and a Mo active site, with overall dimensions of 92× 78×70 Å^3 and an accessible surface area of 38 035 Å^2. The overall structure of the protein (Figure 3.11) is similar to other structurally characterized members of the XO family with a RMSD of 2.1 Å, 2.2 Å, and 2.3 Å for the superposition with *Ta*HBCR[238], *Bt*XO[237], and human aldehyde oxidase (*Hs*AOX1)[162], respectively. Sequence identity between PaoABC and other similar proteins is represented in Figure 3.12.

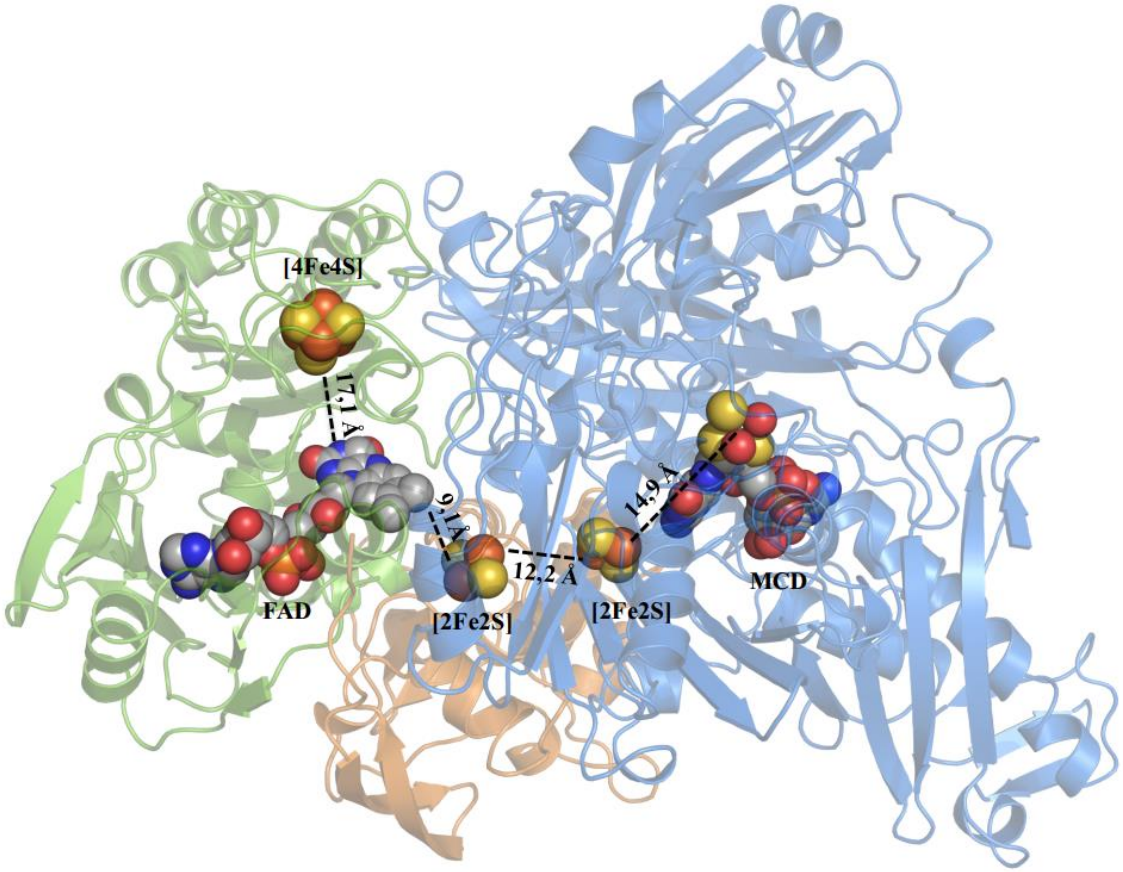


Figure 3.11. Crystal structure of *E. coli* PaoABC. The cofactors are displayed as spheres and are color-coded as atom types.

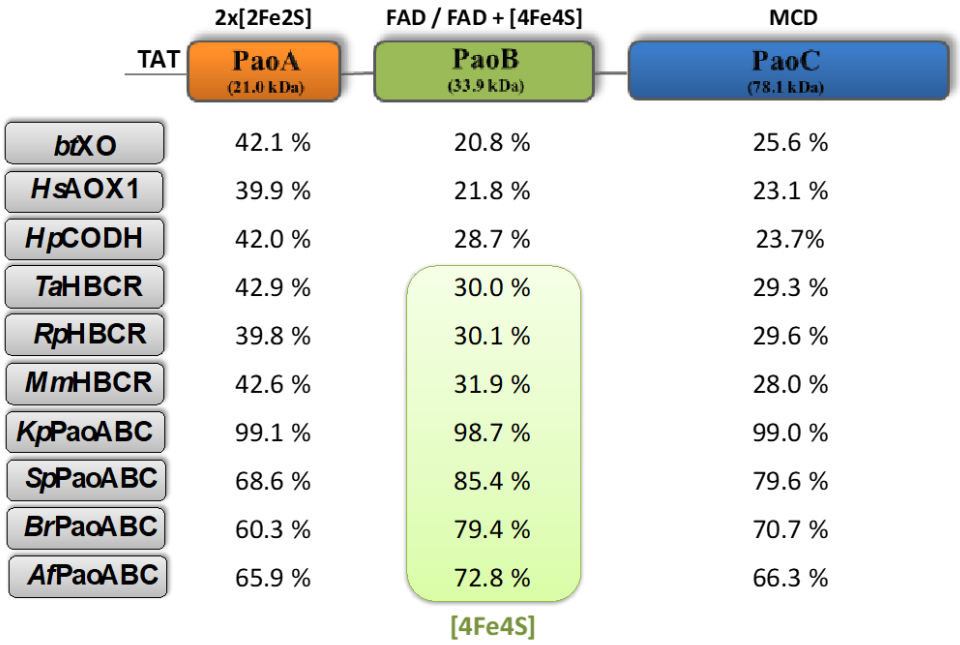


Figure 3.12. Percentage identity between the three subunits of PaoABC and the corresponding subunit of several enzymes from the xanthine oxidase family. Abbreviation: *Bt*XO, *Bos Taurus* xanthine oxidase (PDB: 1fiq); *Hs*Aox1, Homo sapiens aldehyde oxidase (PDB: 4uhw); *Hp*CODH, *Hydrogenophaga pseudoflava* carbon monoxide dehydrogenase (PDB: 1ffv); TaHBCR, *Thauera aromatica* 4-hydroxybenzoyl-CoA reductase (PDB: 1rm6); *Rp*HBCR, *Rhodopseudomonas palustris* 4-hydroxybenzoyl-CoA reductase (UniProtKB: Q21AK5, Q21AK7, Q21AK6); *Mm*HBCR, *Magnetospirillum magneticum* 4-hydroxybenzoyl-CoA reductase (UniProtKB: Q2W5Q2, Q2W5Q0, Q2W5Q1). *Kp*PaoABC, *Klebsiella pneumoniae* aldehyde oxidoreductase family protein (UniProtKB: CDK71072, CDK71073, KME74099); *Sp*PaoABC, *Sodalis praecaptivus* aldehyde oxidoreductase family protein (GeneBank: AHF78136; AHF78137, AHF78138); *Br*PaoABC, *Bradyrhizobium retamae* aldehyde oxidoreductase (GeneBank: KRR22104, KRR22103, KRR22102); *Af*PaoABC, *Asanoa ferruginea* aldehyde oxidoreductase family protein (GeneBank: KOX54054, KOX54053, KOX54052).

The iron−sulfur subunit (PaoA) of the protein comprises 179 residues (depicted in orange in Figure 3.11), with dimensions of 44×37×32 Å^3 and can be divided into two subdomains, each carrying one [2Fe-2S] cluster. According to previous EPR experiments described by Neumann *et al*[243], the clusters were termed type I and type II as determined for the other structurally characterized homologs. The N-terminal domain (residues 53−134) is similar to plant-type ferredoxins[285] and harbors the type II [2Fe-2S] cluster, which is localized next to the FAD in the PaoB subunit (about 9 Å away). The type I [2Fe-2S] in subunit PaoA (residues 135−226) is deeply buried about 19 Å from the protein surface (Figure 3.11).

The PaoC subunit (in blue in Figure 3.11) comprises 729 residues, contains the Moco active site, and has overall dimensions of 75×67×48 Å^3. The active site is analyzed in detail in section 3.3.2.3.

The PaoB subunit comprises 316 residues and has dimensions of 51×48×32 Å^3 (in green in Figure 3.11). It exhibits a typical FAD-binding motif conserved within the XO family and near to the isoalloxazine moiety, toward the solvent, the structure is very similar to that in BtXDH, *Hs*AOX1, and *Ta*HBCR. There is no deviation of the FAD variable loop (loop 430−440) in PaoABC that adopts a different conformation in the XDH to XO interconversion [162,237].

During structure determination, the presence of an unexpected [4Fe-4S] cluster was identified in the protein. Comparison with homologous proteins showed that this cluster corresponds to the [4Fe-4S] center present in the crystal structure of *Ta*HBCR[238]. The [4Fe-4S] cluster of PaoABC is embedded in a 43-residue-long polypeptide segment (PaoB-C119 to PaoB-H161) positioned close to the si-face of FAD. Multiple-sequence alignments gave the highest homology scores for those Moco enzymes that also contain this insertion segment and the cysteine residues responsible for the binding of the [4Fe-4S] cluster (e.g. PaoABC from *Shigella boydii* and *Klebsiella pneumonaie* with 99% sequence identity; Figure 3.12). Most of these enzymes are classified as XDH and only a few as periplasmic aldehyde oxidases or HBCRs. These findings suggest that the presence of a [4Fe-4S] cluster in members of the XO family of enzymes is more common than thought before and may be important for their respective physiological activity. Currently, the mode of function for most of the enzymes mentioned above is not fully clear, except for the *Ta*HBCRs and CODH[246].

The three subunits of PaoABC form a stable heterotrimer, unlike the other members of the XO family so far structurally characterized. These are organized as dimers of heterotrimers $(\alpha\beta\gamma)_2$ (e.g., *Ta*HBCR, CODH) or as homodimers α_2 (e.g., *Bt*XOR, *Hs*AOX1). In the case of XO and AOX, the dimer interface is predominantly between the two catalytic subunits, whereas in *Ta*HBCR, both the catalytic and the 2Fe-2S subunits mediate the dimer formation. Ionic interactions are essential for dimerization (e.g., R802(α)/E765(α') and E768(α)/R801(α') in *Hs*AOX1, R793(α)/E756(α') and E759(α)/K792(α') in *Bt*XO). However, the corresponding residues are not conserved in PaoABC.

It was reported previously for *R. capsulatus* XDH that dimerization is required for Moco insertion[239,253]. In general, proteins form dimers ensuring a higher stability for the protein. In the case of bacterial molybdoenzymes, there seems to be a difference in the oligomerization state depending on the cellular localization of the protein. While cytoplasmic molybdoenzymes, like the bacterial XDH, require formation of dimers via the catalytic subunit for Moco insertion, it does not seem to be necessary for all periplasmic enzymes. Here, the periplasmic enzymes often form multimers of two different subunits, with no dimerization via the Moco subunit.

The biochemical analysis suggested that PaoABC does not dimerize via its Moco-binding domain and remains an heterotrimer ($\alpha\beta\gamma$), in solution. Although only one molecule is found in the asymmetric unit (see Table 3.5, section 3.3.1.1), the crystal packing from the position of the crystallographic dyads were analyzed. This was first done by visual inspection and no tight packing could be identified (Figure 3.13).

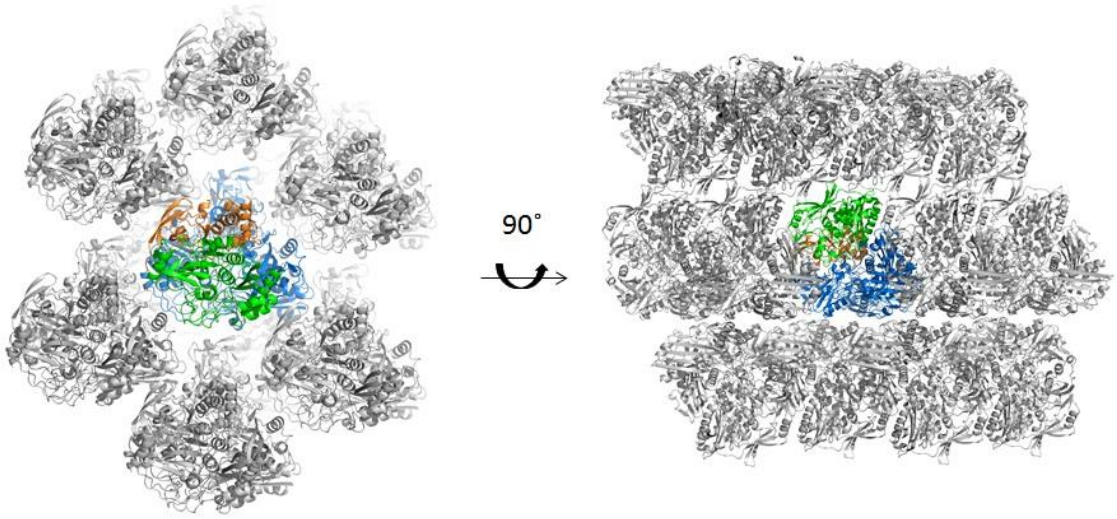


Figure 3.13. Crystal packing of PaoABC. Cartoon representation of several asymmetric units of PaoABC untwinned crystals with subunit A in orange, B in green and C in blue when viewed along the c axis and rotated 90°. Picture prepared using PyMOL.

In addition, the PISA server (Protein Interfaces, Surfaces, and Assemblies)[286] was used and predicted the 'ABC' arrangement based on chemical thermodynamic calculations. The solvation free energy gain (ΔG_{int}) of -163.1 kcal/mol and the free energy of assembly dissociation (ΔG_{diss}) of 22.8 kcal/mol shows that PaoABC is thermodynamically stable as $\alpha\beta\gamma$ heterotrimer.

Other examples of monomers include the formate dehydrogenase ($\alpha\beta$) from *Desulfovibrio gigas*, the formate dehydrogenase 2 ($\alpha\beta\gamma$) from *Desulfovibrio vulgaris*, the sulfite dehydrogenase ($\alpha\beta$) from *Starkeya novella*, the TMAO reductase (α) from *E. coli*, and the DMSO reductase ($\alpha\beta\gamma$) from *E. coli*. Explanations for the lack of dimerization might be that either structural stabilization by dimerization of these proteins is not required in the periplasmic environment or that the translocation via the Tat-translocase has an impact on the oligomerization state. In the periplasmic enzymes, Moco insertion occurs before the translocation to the periplasm and includes a "proofreading" mechanism which ensures that only the fully matured protein is transported in its folded state after the cofactor insertion.

The catalytic subunit of PaoABC has a heart-like shape similar to the other family members that can be further divided into two subdomains running almost perpendicular to each other. Analysis of the superposition of members of the XO family suggest that the N-terminal subdomain of the subunit is highly conserved, especially in the five-stranded β sheets, while some divergence is observed in the C-terminal subdomain, especially in the two loops that are connecting strands β20 to β21 and α16 to β28, designated as loop 1 and loop 2, respectively (Figure 3.14). Sequence alignment of this region shows that these loops are shorter in PaoABC and align with the closely related molybdopterin enzymes that contain the [4Fe-4S] cluster. Loops 1 and 2 correspond to a cap at the surface of the protein that controls the solvent exposure of the active site. While in most Mo enzymes, the cap protects the catalytic site from the solvent, with the metal deeply buried at the bottom of a 10−14-Å long funnel, the same is not true for PaoABC. In this case, the cap is

much shorter and the Mo active site is very exposed to the solvent. The substrate channel commonly observed in the structurally characterized enzymes is absent in PaoABC. The active site is in a shallow groove, very close to the surface of the protein, unlike in *Ta*HBCR where the long and narrow channel accounts for its high substrate specificity.

The easy access and the absence of aromatic residues lining the active site make this protein unique in the XO family of enzymes.

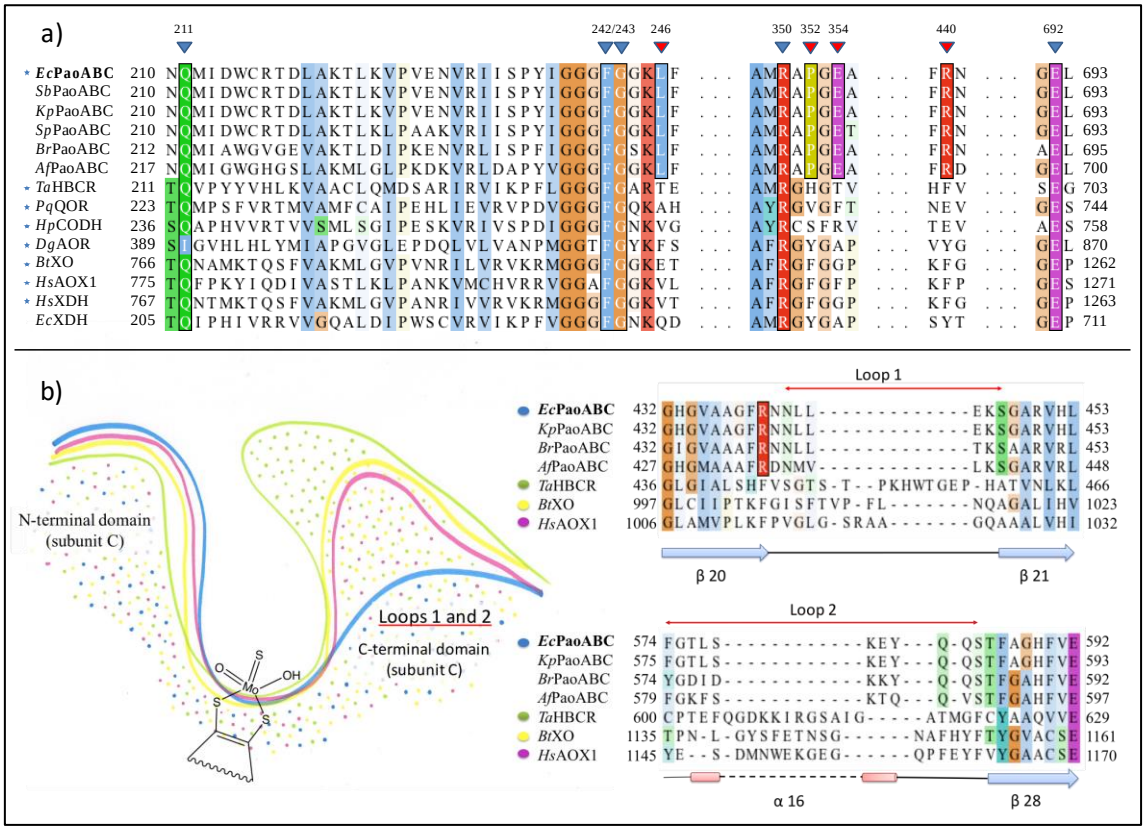


Figure 3.14. **a)** Sequence alignment of the Moco domain of fourteen bacterial members of the molybdenum hydroxylase family. Sequences were aligned using the Clustal Omega. The blue arrows represent the most important conserved amino acids that are involved in the coordination of the MCD, the red arrows the non-conserved amino acids and the star symbol (★) the proteins with x-ray structure deposited on PDB database. Abbreviations: *Ec*PaoABC, *E. coli* periplasmic aldehyde oxidoreductase C subunit; *Sb*PaoABC, *Shigella boydii m*olybdopterin-binding domain of aldehyde oxidoreductase family protein (UniParcKB: UPI00066A1DF4); *Kp*PaoABC, *Klebsiella pneumoniae* molybdopterin-binding domain of aldehyde oxidoreductase family protein (UniProtKB: KME74099); *Sp*PaoABC, *Sodalis praecaptivus* molybdopterin-binding domain of aldehyde oxidoreductase family protein (GeneBank: AHF78138); *Br*PaoABC, *Bradyrhizobium retamae* molybdopterin-binding domain of aldehyde oxidoreductase family protein (GeneBank: KRR22102); *Af*PaoABC , *Asanoa ferruginea m*olybdopterin-binding domain of aldehyde oxidoreductase family protein (GeneBank: KOX54052); TaHBCR, *Thauera aromatica* 4-hydroxybenzoyl-CoA reductase A subunit (PDB: 1rm6); *Pp*Qor subB, *Pseudomonas putida* quinoline 2-oxidoreductase B subunit (PDB:1t3q); *Hp*CODH subB, *Hydrogenophaga pseudoflava* carbon monoxide dehydrogenase B subunit (PDB: 1ffv); *Dg*AOR subA, *Desulfovibrio gigas* aldehyde oxidoreductase Moco domain (PDB: 1vlb); *Bt*XO, *Bos Taurus* xanthine oxidase C subunit (PDB: 1fiq); *Hs*AOX1B, *Homo sapiens* aldehyde oxidase (PDB: 4uhw); *Hs*XDH, *Homo sapiens* xanthine oxidase (oxidoreductase; XOR) FAD domain (PDB: 2e1q); *Ec*XDH, *E. coli* xanthine oxidase FAD subunit (UniProtKB: Q46800). **b)** Scheme of the superposition of the Moco domain from PaoABC (blue), *Hs*AOX1 (pink), TaHBCR (green), *Bt*XO (orange). PaoABC is present in a more open state with the active site easily accessible to the solvent while *Ta*4HBCR and *Bt*XO are present in a more closed state.

SAXS Results

SAXS experiments have also been performed for PaoABC to further explore the structure and oligomerization state of the protein in solution. Monodisperse concentrated solutions of PaoABC were measured and the processed scattering profile of PaoABC are presented in Figure 3.15. The structural parameters including the radius of gyration (R_g), the maximum particle dimension (D_{max}) and the excluded (V_p) volume of the hydrated particle computed from the experimental data are summarized in Table 3.7.

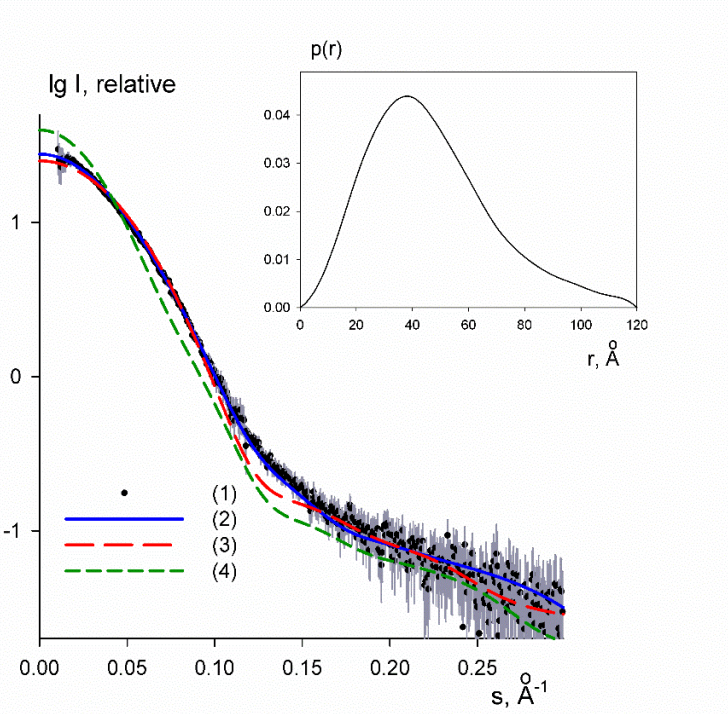


Figure 3.15. SAXS data from PaoABC in solution. The experimental data (1) are displayed as black dots with grey error bars, and the scattering computed from the models is shown as smooth lines: (2) is the scattering from the *ab initio* shape, (3) and (4) are scattering computed from the monomer and dimer of *Ta*HBCR (PDB code 1rm6), respectively. Insert, the distance distribution function of PaoABC computed from the scattering data.

Table 3.7. SAXS Data collection and derived parameters for PaoABC. Values for the data merged from the data collected for 0.3 and 10 mg/mL.

Structural parameters	
R_g, Å [from $P(r)$]	36 ± 1
R_g, Å (from Guinier)	35 ± 1
D_{max}, Å	120 ± 10
Porod volume estimate V_p, Å^3 × 10^3	200 ± 10
Molecular mass determination (kDa)	
From $I(0)$[a]	130 ± 15
From Porod volume[b]	125 ± 10
From sequence	136

[a] Using BSA as a reference. [b] Calculated by $MM = V_p/1.6$

The distance distribution function, $P(r)$, obtained from the experimental data suggests that the protein is well folded and has a globular structure (Figure 3.15). The molecular mass of the solute estimated from the forward scattering $I(0)$ and from the Porod volume (Table 3.7) are both close to that of the monomeric PaoABC (136 kDa) indicating that the protein is a heterotrimer in solution. The overall shape of PaoABC was calculated *ab initio* from its scattering profile using programs DAMMIN[73] and DAMMIF[91]. The most typical model out of 20 reconstructions as analyzed by DAMAVER[75] is overlapped in Figure 3.16 with the crystallographic structure of the monomer of the *Ta*HBCR (PDB code 1rm6 [238]). The crystal structure of PaoABC was not used because, at that time, the crystallization was under optimization. The reconstructed shape matches well with the overall appearance of the crystal structure of the homologue. The experimental SAXS data were also fitted by the scattering profile calculated from the crystal structure of the monomeric 4-hydroxybenzoyl-Coa reductase using CRYSOL[94]. The crystal structure yielded R_g of 31 Å, somewhat smaller than the experimental value. The fit showed a discrepancy of χ^2 = 1.3 (Figure 3.15) confirming that the overall shape of PaoABC is reasonably close to that of monomeric *Ta*HBCR, but also showing systematic deviations and indicating that PaoABC may be somewhat more extended in solution. In contrast, the scattering computed from the crystallographic dimer of *Ta*HBCR provides an extremely poor fit to the SAXS data (Figure 3.15) with discrepancy χ^2 = 3.1 and the radius of gyration of the dimer (R_g = 43 Å) does not match the experimental data. Taken together, the SAXS data indicate that PaoABC is a heterotrimeric protein in solution with the overall structure similar but somewhat more extended than that of the monomeric 4-hydroxybenzoyl-Coa reductase.

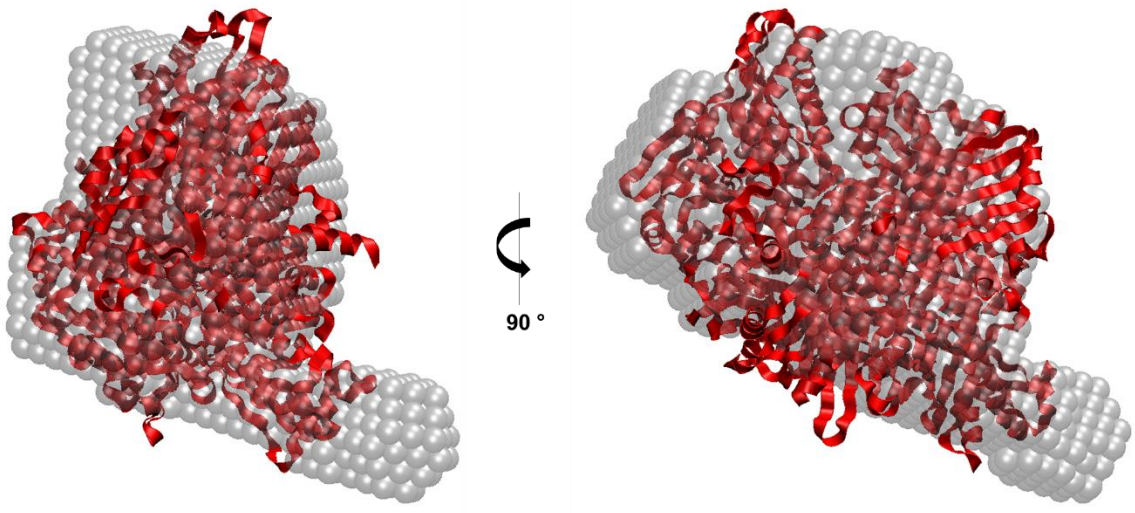


Figure 3.16. Superposition of the *ab initio* envelope of PaoABC with a homologous structure. Transparent beads – *ab initio* model; cartoon representation of the *Ta*HBCR monomer (PDB code 1rm6).

3.3.2.2. The unexpected [4Fe-4S] cluster

The discovery of an extra cofactor in PaoB, a [4Fe-4S] cluster, was very surprising since previous studies did not anticipate this result[243]. The cluster is buried approximately 12 Å beneath the protein surface and is coordinated by cysteines 119, 129, 138, and 157 that belong to a 43 residue insert in the polypeptide chain. This insertion corresponds to a coiled region with no secondary structure elements except a short α-helix that is not structurally related to *Ta*HBCR (the RMSD value for the superposition of 38 Cα atoms is 6.1 Å).

In PaoABC, the [4Fe-4S] cluster is *ca*. 17 Å distant from the isoalloxazine ring of FAD, which is a large distance for an electron transfer process. Nevertheless, as found in *Ta*HBCR, key residues may mediate the putative electron path, connecting the [4Fe-4S] cluster and the remaining redox centers toward the Mo active site. As depicted in Figure 3.17, PaoB-R118 and PaoB-F239 are in close distance to the [4Fe- 4S] cluster and the FAD, respectively, and may facilitate a possible electron transfer route between the two centers. For the case of *Ta*HBCR, the equivalent residues (R121 and F233, 4-HCBR numbering) were considered to be important for the reductase activity of the enzyme, contributing to the electron flux from the [4Fe-4S] center to the substrate [238]. PaoABC has been classified as a detoxifying enzyme in the metabolism of aldehydes to less toxic carboxylic acids, catalyzing the hydroxylation reaction of aromatic aldehydes. Our collaborators have strong experimental evidence that PaoABC is not able to perform the reverse reaction, the reduction of carboxylic acids into aldehydes, and so far, the involvement of the [4Fe-4S] cluster during catalysis is not clear. One possible involvement of the [4Fe-4S] cluster might be to prevent the formation of a flavin semiquinone at the FAD site since the formation of the semiquinone has never been detected during the enzyme reduction [243]. A stable protein could not be obtained when mutants with one of the four coordinating cysteines (e.g., mutations at PaoB-C138 or PaoB-C157) were made, indicating that the [4Fe-4S] cluster is required to maintain the integral fold of the PaoB subunit.

The amino acid sequence of PaoB was used as query to search for other enzymes containing a FAD-binding subunit with four cysteines binding motif ($CX_9CX_7CX_{18}C$). It appears that the existence of subunits harboring FAD and the cysteine binding motif is common in bacterial strains even though some deviation in the number of residues separating the Cys residues is observed, as expected. Based on a BLAST search, these subunits belong to Mo-containing enzymes predominantly found in Gram-negative proteobacteria in all its subdivisions and also in other bacteria phyla such as Green Non-Sulfur bacteria (e.g., *Ktedonobacter racemifer*, 51% identity), nitrogen-fixing bacteria (*Bradyrhizobium retamae*, 80% identity), cyanobacteria (e.g., *Tolypothrix campylonemoides*, 53% identity), verrucomicrobia (e.g., *Chthoniobacter flavus*, 52% identity), or in Gram positive species from firmicutes (e.g., *Paenibacillus sophorae*, 53% identity), or actinobacteria (e.g., *Asanoa ferruginea* and *Paenibacillus sophorae*, 73 and 53% identity, respectively). Considering up to 98% of sequence coverage, 54 hits were found with over 80% identity, after excluding all *Escherichia* species from the search. For most of the enzymes, a high degree of sequence identity is also found in the Mo containing subunit (eg *Shigella boydii*,

Klebsiella pneumonaie, *Bradyrhizobium retamae*, and *Asanoa ferruginea* with 99%, 99%, 71%, and 66% identity, respectively). To further characterize the phylogenetic relationships between these enzymes, a dendrogram was created with Phylogeny.fr[31,32] using PaoB-related sequences from 67 bacterial strains, belonging to the five phyla. The sequence of the FAD containing subunit of XDH from *E. coli* and from *Ta*HBCR was also included as negative controls since the first does not have a [4Fe-4S] cluster and the latter, although containing such a cluster, catalyzes a reduction reaction unrelated to aldehyde oxidation. The resulting unrooted tree is shown in Figure A2, Appendix. Bootstrap analysis was also performed, and the obtained values are shown in the branch points of the tree. This analysis suggests that the function of the periplasmic enzymes containing the [4Fe-4S] cluster is markedly different from XDHs and HBCRs, which appear in a distant and separate branch of the dendrogram. In addition, this dendrogram suggests horizontal gene transfer events between unrelated species from five different bacteria phyla. Most of the enzymes used are classified as periplasmic with the corresponding Moco subunit displaying 35% to 72% sequence identity to PaoC. These findings point to the existence of an extensive number of XO-type enzymes harboring an additional [4Fe-4S] cluster. Thus, we propose that these enzymes will have a similar function and subcellular localization as PaoABC and will be involved in the periplasmic oxidation of aldehydes and acting as detoxifying enzymes.

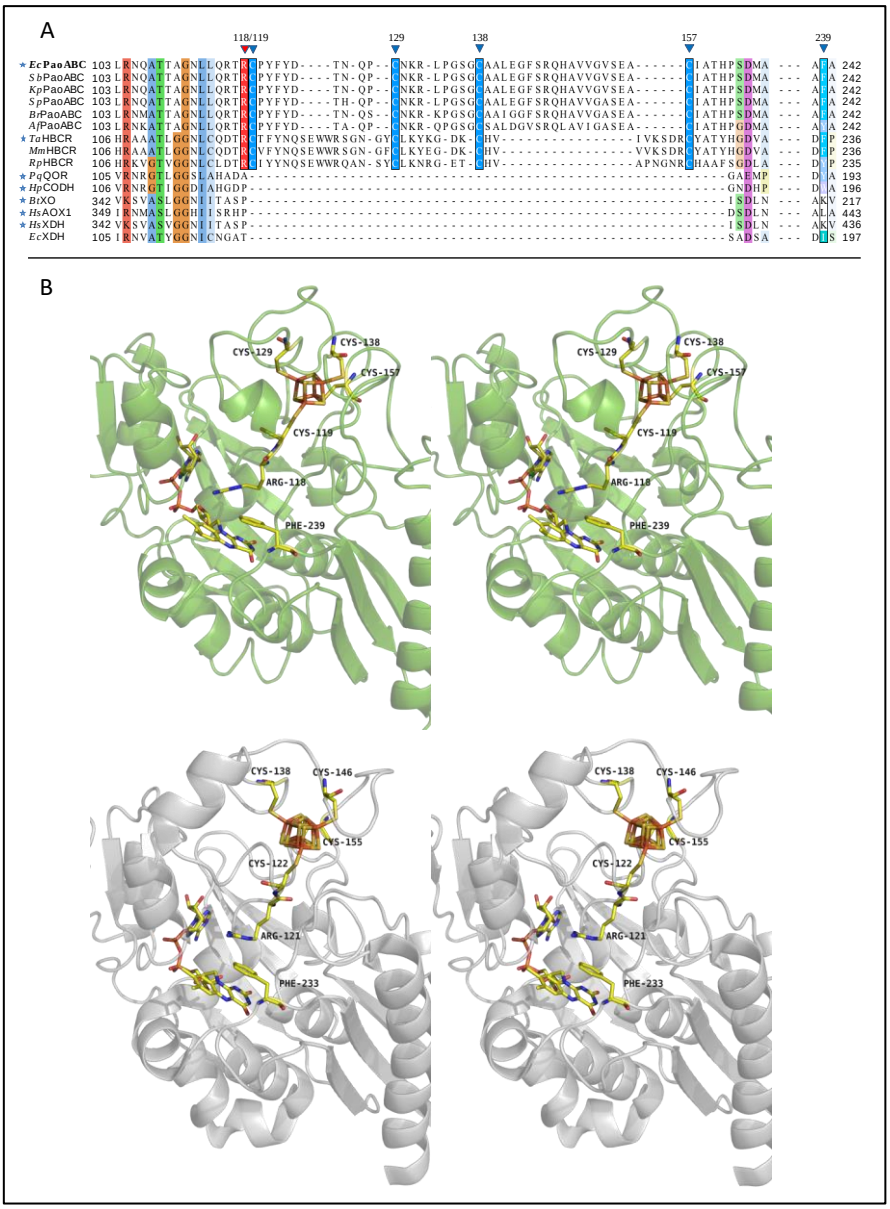


Figure 3.17. A. Sequence alignment of the FAD domain of 16 bacterial members of the molybdenum hydroxylase family. Sequences were aligned using the Clustal omega. The blue arrows represent the four cysteines that coordinate the [4Fe-4S] cluster; the red arrows, the R118 and F234 that are probably involved in the electronic path from FAD to the [4Fe-4S] cluster; and the star symbol (★), the proteins with X-ray structure deposited in the PDB. Abbreviations: EcPaoABC, *E. coli* periplasmic aldehyde oxidoreductase B subunit; SbPaoABC, *Shigella boydii* aldehyde oxidoreductase family protein FAD subunit (UniParcKB: UPI00066BBD47); KpPaoABC, *Klebsiella pneumoniae* aldehyde oxidoreductase family protein FAD subunit (GeneBank: KME74100); SpPaoABC, *Sodalis praecaptivus* aldehyde oxidoreductase family protein FAD subunit (UniProtKB: AHF78137.1); BrPaoABC, *Bradyrhizobium retamae* aldehyde oxidoreductase FAD subunit (UniProtKB: A0A0R3MPP7); Af PaoABC, *Asanoa ferruginea* aldehyde oxidoreductase family protein FAD subunit (GeneBank: KOX54053); TaHBCR, *Thauera aromatica* 4-hydroxybenzoyl-CoA reductase B subunit (PDB: 1rm6); MmHBCRB, *Magnetospirillum magneticum* 4-hydroxybenzoyl-CoA reductase B subunit (UniProtKB: Q2W5Q0); RpHBCRB, *Rhodopseudomonas palustris* 4-hydroxybenzoyl-CoA reductase B subunit (UniProtKB: Q21AK7); PpQor, *Pseudomonas putida* quinoline 2- oxidoreductase C subunit (PDB: 1t3q); HpCODH, *Hydrogenophaga pseudoflava* carbon monoxide dehydrogenase (CODH) C subunit (PDB: 1ffv); BtXO, *Bos taurus* xanthine oxidase B subunit (PDB: 1fiq); HsAOX1, Homo sapiens aldehyde oxidase (PDB: 4uhw); HsXDH, *Homo sapiens* xanthine oxidase (oxidoreductase; XOR) FAD domain (PDB: 2e1q); EcXDH, *E. coli* xanthine oxidase FAD subunit (UniProtKB: Q46800). **B.** Stereo representation of the insertion segment of the [4Fe-4S] center domain for PaoABC (green) and *Ta*HBCR (gray). The insertion segment is composed of 43 amino acids from Cys(B)119 to His(B)161 and wraps the [4Fe-4S] cluster. The distances between the irons and the cysteine sulfur atoms vary between 2.2 and 2.4 Å.

3.3.2.3. Active site

PaoABC has been characterized as the only example of an *E. coli* Moco enzyme where cytosine is found in the dinucleotide form of the cofactor. In the present crystal structure, MCD is very well-defined in the electron density maps, where a net of hydrogen bonds contributes to the stabilization of the entire cofactor. Like in most Moco enzymes deposited in the PDB, the tricyclic pyranopterin ring system forces a conserved arginine residue in the catalytic domain, PaoC-R350, to adopt an unfavorable conformation that is easily spotted in the Ramachandran plot, bringing its guanidinium moiety coplanar with the pterin aromatic system.

In PaoABC, the Mo atom is coordinated to the dithiolene moiety, to an apical oxo ligand, an equatorial sulfido ligand, and a labile hydroxo group in a distorted square pyramidal geometry, similar to most of the enzymes of the XO family. Unlike other ligands of the metal, a high B factor was observed for the sulfur ligand of the equatorial position. This discrepancy led us to suspect a low occupancy of the sulfido ligand or its replacement by an oxo group, which would correspond to the desulfo form of the protein. Considering the cyanolyzable content of the protein (58%), previously reported [243], the sulfido ligand of PaoABC has been modeled with 60% occupancy, and the corresponding B factor is in the same range as the other ligands (12 $Å^2$).

Analysis of active site contacts shows the Mo equatorial hydroxyl ligand within hydrogen-bonding distance (2.9 Å) to the backbone nitrogen of PaoC-G508 (for clarity, this residue is not shown in Figure 3.18) while the apical oxo-group (OM1) is hydrogen-bonded to the Nε2 of the highly conserved PaoC-Q211 (2.9 Å). Moreover, PaoC-E692, implicated in the catalytic reaction mechanism of all XO-related enzymes, occupies the same position as in the other structurally characterized enzymes (Figures 3.15 and 3.17). Site-directed mutagenesis studies have already been performed for this residue, and the exchange of E692 for glutamine resulted in an inactive enzyme[243].

Despite the similarities with other enzymes from the XO family, the active site of PaoABC is remarkably different. The presence of a proline (PaoC-P352) occupying a similar position where a phenylalanine (F923/F914 in *Hs*AOX1/*Bt*XO) or a histidine (H360 in *Ta*HBCR) are found is unexpected (Figures 3.14, 3.18 and Table 3.8). Those planar residues are often involved in stacking interactions with the substrate molecules.

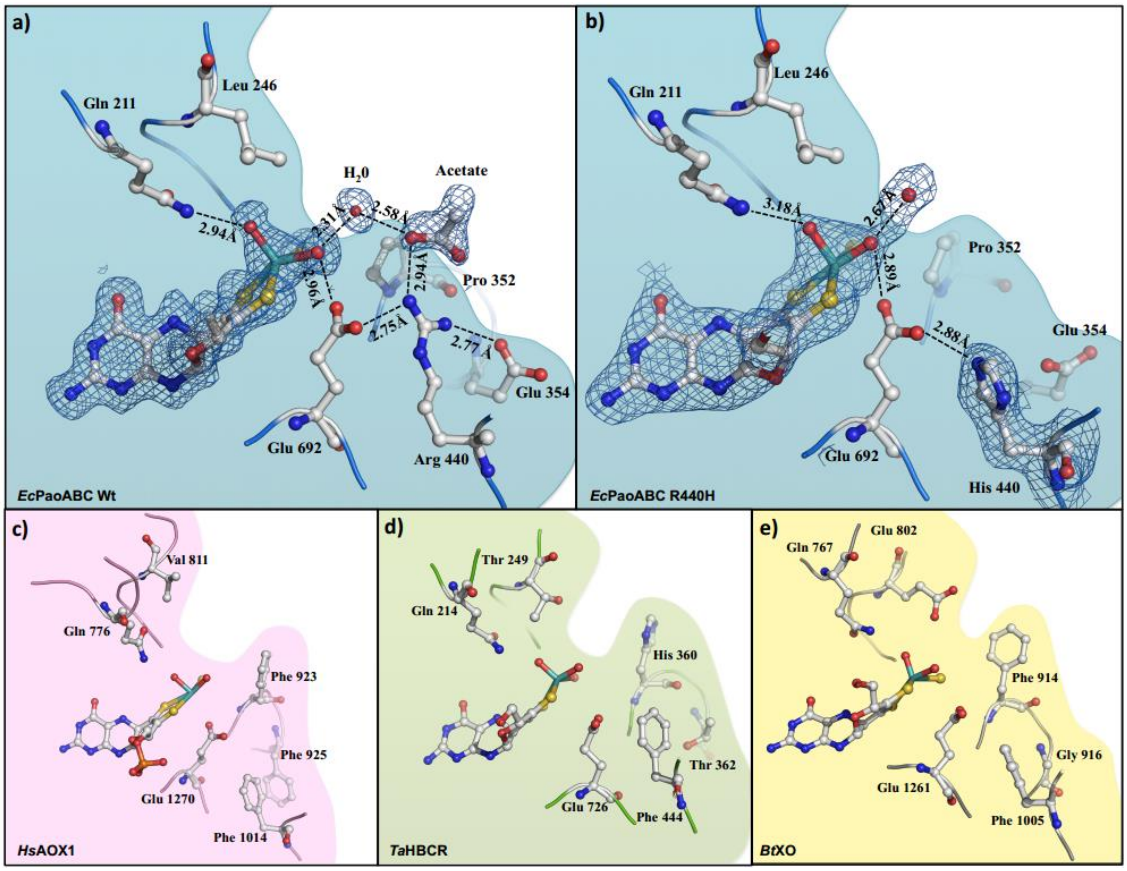
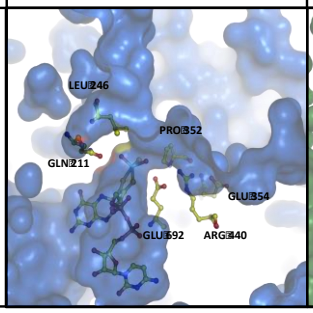
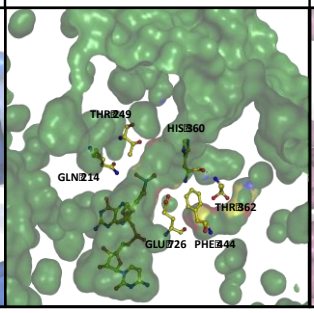
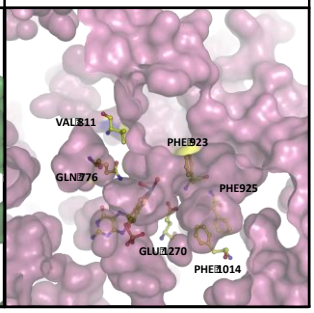


Figure 3.18. a) The Mo active site of *Ec*PaoABC for the wild-type, **b)** *Ec*PaoABC R440H mutant, **c)** *Hs*AOX1, **d)** *Ta*HBCR, **e)** *Bt*XO. The electron density map at a contour level of 1.0 σ is shown in blue. Moco and selected amino acids are displayed as balls-and-sticks and are atom color-coded.

Also notable is PaoC-R440 positioned close to the catalytic PaoC-E692. The side chain of PaoC-R440 is replacing the aromatic residues typical of related enzymes (F1014/F1005 in *Hs*AOX1/*Bt*XOs; Figure 3.18). In fact, there are no aromatic side chains that could be responsible for substrate stabilization during catalysis in the active site of PaoABC. The only aromatic residues near to the cofactor are at *ca.* 7−9 Å from the Mo (PaoC-H187, F247, W215, and W510), and no stacking interactions are expected. PaoC-R440 is at 2.77 Å from PaoC-E354, and these two residues are highly conserved in other putative Moco enzymes with the [4Fe-4S] binding motif at the FAD subunit (Figures 3.14, Figure 3.17 and Table 3.8).

To understand the importance of PaoC-R440 in the reaction mechanism, site-directed mutagenesis was carried out, and two mutants of PaoABC, PaoC-R440H and PaoC-R440K, were produced. The mutants were successfully expressed and purified and used for kinetic studies[287] and crystallization. Kinetic studies showed that the PaoABC mutant R440H have a slightly increased on K_M and a small decrease on K_{cat} (395.9 µM and 293.0 s^{-1}, respectively) when compared to the wild-type (K_M = 306.7 µM and K_{cat} = 486.0 s^{-1}). For wild-type and mutants, the activity with benzaldehyde was detectable between pH 4.0 and pH 5.5, which is in accordance with previous reports [243,251].

Table 3.8. Main features of *Escherichia coli* PaoABC, *Thauera aromatica* 4-hydroxybenzoyl-CoA reductase (*Ta*HBCR) and *Homo sapiens* aldehyde oxidase (*Hs*AOX1).

	***Ec*PaoABC**	***Ta*HBCR**	***Hs*AOX1**
Cellular localization	Periplasm	Cytoplasm	Cytoplasm
Subunit composition	A – 2x[2Fe2S] B – FAD + [4Fe4S] C – MCD	A – MCD B – FAD + [4Fe4S] C – 2x[2Fe2S]	A – 2x[2Fe2S] + FAD + MPT
Reaction catalyzed	Cinnamaldehyde + H_2O → cinnamic acid+ $2H^+ + 2e^-$	4-OHBenzoylCoA + $2H^+ + 2e^-$ → BenzoylCoA + H_2O	RH + H_2O → ROH + $2H^+ + 2e^-$
Conserved active site residues	Gln(C)211 / Glu(C)692	Gln(A)214 / Glu(A)726	Gln776 / Glu1270
Non-conserved active site residues	Leu(C)246 / Pro(C)352 / Glu(C)354 / Arg(C)440	Thr(A)249 / His(A)360 / Thr(A)362 / Phe(A)444	Val811 / Phe923 / Phe925 / Phe1014
Substrate channel			

The crystal structure of the PaoC-R440H mutant was also determined to 2.3 Å resolution, and the refined model is almost identical to the wild-type structure (RMSD of 0.17 Å for the superposition of 1212 Cα atoms). Clear electron density was found for the imidazole side chain (H440), which is hydrogen bonded to the PaoC-E692 at 2.88 Å (Figure 3.18). In the crystal structure of this mutant, the side chain of PaoC-E354 is in a different position compared to the wild-type, which may suggest the involvement of this conserved glutamate in the reaction mechanism.

Furthermore, additional electron density was found close to the active site of the wild-type, filling part of the substrate groove. This can be interpreted and refined as an acetate ion, probably from the buffer solution. Other possible anions such as bicarbonate cannot be excluded. This negatively charged ion is 2.94 Å from the positively charged PaoC-R440 and interacting with the labile hydroxyl ligand of the active site through hydrogen bonds mediated by a water molecule (Figure 3.18). The position and orientation of the ion suggest the position of the substrate in the active site during catalysis, and in fact, this anion is absented in the crystal structure of the PaoC-R440H mutant.

To characterize the interaction between PaoABC and its physiological substrates (cinnamaldehyde, benzaldehyde, and others) and also with compounds biotechnological interest (such as 2,5-diformylfuran and 5-formylfuran-2-carboxylic acid[288]), several attempts have been carried out to obtain the crystal structure of a protein−ligand complex using substrate analogs or putative inhibitors. However, structures of complexes have not been obtained yet, and all attempts resulted in the ligand-free form of PaoABC. The presence of a negatively charged ion putatively blocking the access to the active site and/or its high degree of solvent exposure might be responsible for the difficulties in obtaining the structure of a ligand-bound form of the enzyme.

Chapter 4

Structural characterization of a Förster resonance energy transfer (FRET)-based molecular beacon using SAXS

Part of the work described in this chapter was the subject of one publication, currently under evaluation:

- Cordeiro M*****, **Otrelo-Cardoso AR***, Svergun DI, Konarev PV, Lima JC, Santos-Silva T, Baptista PV. Light encoded DNA biosensor coupled to SAXS based structural interpretation. *****These authors contributed equally to this work.

The structural characterization of a DNA biosensor based on the Förster resonance energy transfer (FRET) phenomenon was proposed using SAXS. This work, although not related to the main subject of the Thesis, allowed application of the previously described methodologies for protein samples in a different context, namely using nucleic acids.

4.1. General concepts

Small-angle X-ray scattering is a powerful tool to study the dynamic and structural features of complex biological ensembles, such as protein-protein, protein-nucleic acid or nucleic acids-nucleic acids (for more detail see Chapter 1.3.). Due to the electron-rich phosphate backbone of the nucleic acids, they present a high contrast in scattering experiments[68]. Traditional approaches to characterize nucleic acid ensembles are usually performed in a dried sample by atomic force microscopy or transmission electron microscopy or using labeled oligos via fluorescence spectroscopy/ microscopy [289]. SAXS allows the structural characterization of this systems in solution, without prior chemical modification/label (like fluorophores or isotope labeling). The literature is populated with several examples of the application of SAXS in nucleic acids. In one of the examples, Bruetzel *et al* demonstrate that SAXS can quantitatively resolve the conformational changes of a DNA origami two-state switch device[290]. In this study, we explore the combined use of FRET and SAXS for the optical and structural characterization of a two-component FRET-based molecular beacons (MBs). This serves a proof of concept for the applicability of SAXS to short nucleic acid (from 10 to 43bp) molecules[290].

The traditional MB are composed of a single strand DNA molecule (ssDNA) in loop configuration – hairpin (A), recognition element - labeled on one extremity with a fluorophore and with a quencher in the opposite extremity. This recognition element is flanked by auto-complementary sequences (palindromic sequences), that keeps the fluorophore near the quencher, leading to the suppression of the fluorophore emission. This linear ssDNA will exhibit a hairpin structure due to the hybridization of the palindromic sequences, forming a double strand on the extremities of the DNA molecule – stem portion – while the recognition element remains in a single strand and ready to hybridize to its target sequence. Upon positive recognition, the hairpin is disrupted and separates the fluorophore and quencher, allowing for a partial recovery of emission from the fluorophore, rendering a positive signal detection.

The proposed system is composed by a two-component MB (Figure 4.1), where the conventional hairpin structure is used. Here, the hybridization of a complementary sequence to the loop partition of the hairpin leads to its disruption, exposing the palindromic sequence that can be targeted by a specific acceptor labeled oligonucleotide (revelator), (C) [291–294]– Figure 4.1. Hybridization of the revelator brings the donor and acceptor fluorophore in the vicinity of one another, allowing the occurrence of FRET. This generates a specific FRET signature that

corresponds to the detection of the target molecule generating a positive and detectable signal in the presence of the correct revelator.

The design of the traditional MB has been further applied in various systems [295,296] including Au-nanobeacons [297,298].The two-component MB was applied in spectral codification platforms (BioCode) for the detection of fusion sequences associated with the development of chronic myeloid leukemia (CML). Here, the partial recovery of the donor fluorophore emission upon to target recognition is coupled to the hybridization of an acceptor fluorophore-labeled oligonucleotide, generating FRET signals for increased specificity[299]. As such, there is a wide range of sensors platforms whose output signals are dependent on a dynamic structure that responds to the presence/absence of a given analyte. Amongst the plethora of detection schemes, the use of MBs provides simultaneously an enhanced sequence discrimination and signal output via FRET [299–303].

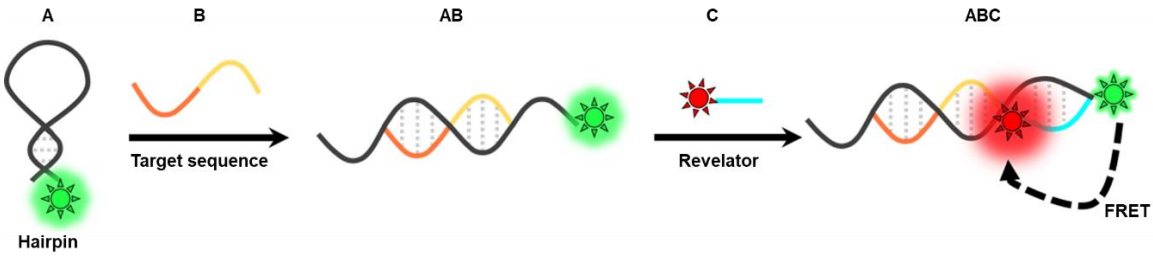


Figure 4.1. Schematic representation of the recognition principle used in the developed biosensor. A – hairpin in the close conformation; B – target sequence; AB – secondary structure disruption due to hybridization to the target sequence; C – revelator/acceptor labeled oligonucleotide; ABC – final structure. Hybridization of acceptor labeled oligonucleotide to exposed hairpin sequence renders a positive FRET signal.

The ssDNA structures were designed by Cordeiro *et al*[299] to generate a hairpin structure, and optimized to specifically detect the fusion of the Abelson oncogene (ABL1) with the breakpoint cluster region (BCR) gene, associated with the e14a2 (Accession number: AJ131466.1) and e13a2 (Accession number: AJ131467.1) transcripts sequences. This feature characterizes the CML that affects nearly 1.5 million people worldwide. Early onset diagnosis coupled to current therapeutics allow for a treatment success rate of 90%, which supports the research on the development of novel diagnostics approaches[304].

In the platforms for molecular recognition proposed [299] for cancer diagnostics, the hairpin and revelator sequence composition depends on the target sequence, e13a2 or e14a2 variants -Table 4.1.

Table 4.1. Oligonucleotide sequences, target specificity and revelators. Adapted from [299].

Biosensor components	Oligonucleotide sequence (5′ → 3′)	nt
Hairpin – e13a2	ccacgccaaacgctgaagggcttcttccttatttttggcgtgg	43
Hairpin – e14a2	cacctcgaaatctgaagggcttttgaactctgttttcgaggtg	43
e13a2	ataaggaagaagcccttcagcg	22
e14a2	cagagttcaaaagcccttcag	21
Revelator - e13a2	ccacgccaaa	10
Revelator - e14a2	cacctcgaaa	10
Partial-complementary sequence – e13	tccgctgaccatcaataaggaagaa	25
Partial-complementary sequence – e14	cactggatttaagcagagttcaaaa	25

4.2. Experimental procedure

The work described in this chapter results from a collaboration with Prof. Pedro V. Baptista, Prof. João Carlos Lima and the MSc. Milton Cordeiro from Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa. The SAXS analysis was performed under the guidance with Prof. Dmitri Svergun and Dr. Peter Konarev from European Molecular Biology Laboratory-Hamburg Outstation, Germany.

All reagents were of analytical grade and purchased from Sigma-Aldrich (Germany). All buffers were filtrated through an Acrodisc® Syringe Filter 0.2 µm Supor® membrane low protein binding Non-Pyrogenic (Pall, United States of America). All oligonucleotides were purchased from STABVIDA (Portugal) and used without further purification. The concentration of all oligonucleotides was determined from the absorbance at 260 nm, measured in an Evolution™ 300 UV-Vis spectrophotometer (Thermofisher, United States of America), using the extinction coefficient provided by the manufacturer.

4.2.1. SAXS data collection and analysis

The SAXS experiments were performed at EMBL P12 beamline, DESY, Hamburg, Germany and at EMBL BM29 beamline, ESRF, Grenoble, France. For the SAXS experiments, the sequences were not labeled with fluorophores. Each component (e13A – hairpin towards the e13a2 fusion sequence, e14A - hairpin towards the e14a2 fusion sequence, e13B - e13a2 fusion sequence, e14B - e14a2 fusion sequence, e13C – revelator for e13A and e14C - revelator for e14A) and each hybridization step (e13AB, e14AB, e14A-e13B-e14C, e13A-e14B-e13C, e13ABC and e14ABC) was measured at 283 K (except e13A, e13B and e13C that were measured at 277 K) in a serial dilution manner, from 2 mg/mL to 0.25 mg/mL in 0.5 × TBE pH 8.3 and 154 mM NaCl – see Table 4.2. The data at P12 beamline was recorded using a Pilatus 2M detector (DECTRIS, Switzerland) with 20 × 0.05 s exposures time, at sample-detector distance 3.00 m and wavelength 1.24 Å. The data at BM29 beamline were recorded using a Pilatus 1M detector with 10 × 1 s exposures time, at sample-detector distance 2.87 m and wavelength 0.99 Å. No measurable radiation damage was detected by comparison of successive time frames.

The primary data reduction was done automatically by the pipeline software, EDNA[191] (for BM29, ESRF) or SASFLOW[81] (in P12, Petra III). The data was processed with the ATSAS package [85] using standard procedures, corrected for buffer contribution, and extrapolated to infinite dilution using PRIMUS[192] from the ATSAS package. GNOM[88] was used to provide the $P(r)$ and determine the corresponding maximum particle size (D_{max}) and the radius of gyration (R_g). The excluded volume of the hydrated DNA molecule (V_P) was calculated using the Porod approximation[69]. The overall shape of each component and each hybridization step were calculated *ab initio* from its

scattering profile using the program DAMMIF[91]. The most typical model out of 20 independent reconstructions was determined by DAMAVER[75] and SUPCOMB[205]. The reconstructed shapes were compared with the overall appearance of the models calculated by NUPACK[305] – Table A2, Appendix.

The collected SAXS data and the generated models have been deposited and are available at SASBDB[194]: SASDC95 (e14A), SASDCA58 (e14B), SASDCB5 (e14C), SASDCC5 (e14AB), SASDCD5 (e14ABC), SASDCE5 (e13A), SASDCF5 (e13B), SASDCG5 (e13C), SASDCH5 (e13AB), SASDCJ5 (e13ABC), SASDCK5 (e13Ae14Be13C) and SASDCL5 (e14Ae13Be14C).

Table 4.2. Different biosensor component analyzed through SAXS and FRET.

	Components	**FRET signal**	**Schematic representation**
Hairpin	e13A e14A	⊖	
Fully complementary target	e13AB e14AB	⊖	
Partial complementary template	e13A-e14B-e13C e14A-e13B-e14C	⊖	
Complementary acceptor	e13ABC e14ABC	⊕	

4.3. Results and discussion

The aim goal of this collaboration was to characterized the several components of a biosensor, the intermediate and final ensemble, using a structural technique such as SAXS and X-ray crystallography. Initially, several attempts were performed to crystallize the hairpin (e13A and e14A), target sequence (e13B and e14B) and revelator (e13C and e14C). Several crystallization conditions were tested but no diffracting crystals were obtained. Difficulties on crystallization could be related with the high flexibility degree of the oligonucleotide sequences. Due to the dynamic character of the systems, SAXS reveals to be a good and fast way for structural characterization, allowing the mimetization of the ideal conditions for detection. Due to the different scattering behavior of the nucleic acids when compared with the proteins, several travels to the synchrotron were necessary to optimize the oligonucleotide concentration.

From the analysis of the scattering profile for the hairpin to detect e13a2 (e13A) and e14a2 (e14A), a R_g of 2.36 nm and 2.93 nm was obtained, respectively – Table 4.3. This value together with the D_{max} are compatible with a close hairpin structure with 43 nt (considering a symmetrical hairpin with a base separation of 3.4 Å). These parameters are slightly higher for hairpin e14A (R_g 2.93 nm and D_{max} 12.5 nm) than for e13A (R_g of 2.36 nm and D_{max} of 8.0 nm) suggesting that the latter adopts a more compact configuration – Table 4.3 and Figure 4.2 (1). This corroborates with *in silico* predictions (using the NUPACK software), where e13A has a higher self-complementarity level than e14A, with a lower free Gibbs energy (- 21.77 kcal/mol) – Table A2, Appendix. Also, the *ab initio* bead models determined for the hairpin e14A and e13A are concordant with the NUPACK simulation – Figure 4.2. At this stage, the calculated models are only interpretation of the SAXS pattern, being necessary to validate them with alternative structural data, as such by X-ray crystallography. The design sequences don't have corresponding atomic data for validation, however the structural analysis corroborates with the basal fluorescence detected in the emission spectrum for the hairpin (Figure A4, Appendix).

The target and revelator were also analyzed individually. For the target sequence e13B, the R_g and D_{max} (1.86 nm and 7.0 nm, respectively) are slightly smaller than for e14B (R_g 2.05 nm and D_{max} 9.0 nm) – Table 4.3 and Figure 4.2. (2). According to the NUPACK prediction, e13B as a tendency to a form a compact structure due to a self-complementary region in the middle of the oligonucleotide sequence justifying these differences - Table A2, Appendix. The compact structure observed in solution does not compromise the hybridization to the hairpin, once they form a more stable interaction. For the revelator sequences, e13C and e14C, the parameters obtain are compatible with a short and linear ssDNA chain – Figure 4.2. (3) and Table 4.3.

Table 4.3. The overall structural parameters estimated from SAXS data. χ^2 - fit quality of *ab initio* models to experimental data.

Sample	R_g nm	D_{max} nm	V_P nm^3	χ^2
e13A	2.36	8.0	24	1.58
e13B	1.86	7.0	13	0.85
e13C	1.32	5.0	8	0.86
e13AB	5.40	20.0	103	1.85
e13ABC	4.56	18.0	59	1.78
e14A	2.93	12.5	34	1.34
e14B	2.05	9.0	15	0.93
e14C	1.29	5.0	7	0.91
e14AB	5.85	22.5	159	1.31
e14ABC	5.21	20.0	105	1.87
e13A-e14B-e13C	3.95	18.0	65	0.76
e14A-e13B-e14C	3.80	14.0	42	1.29

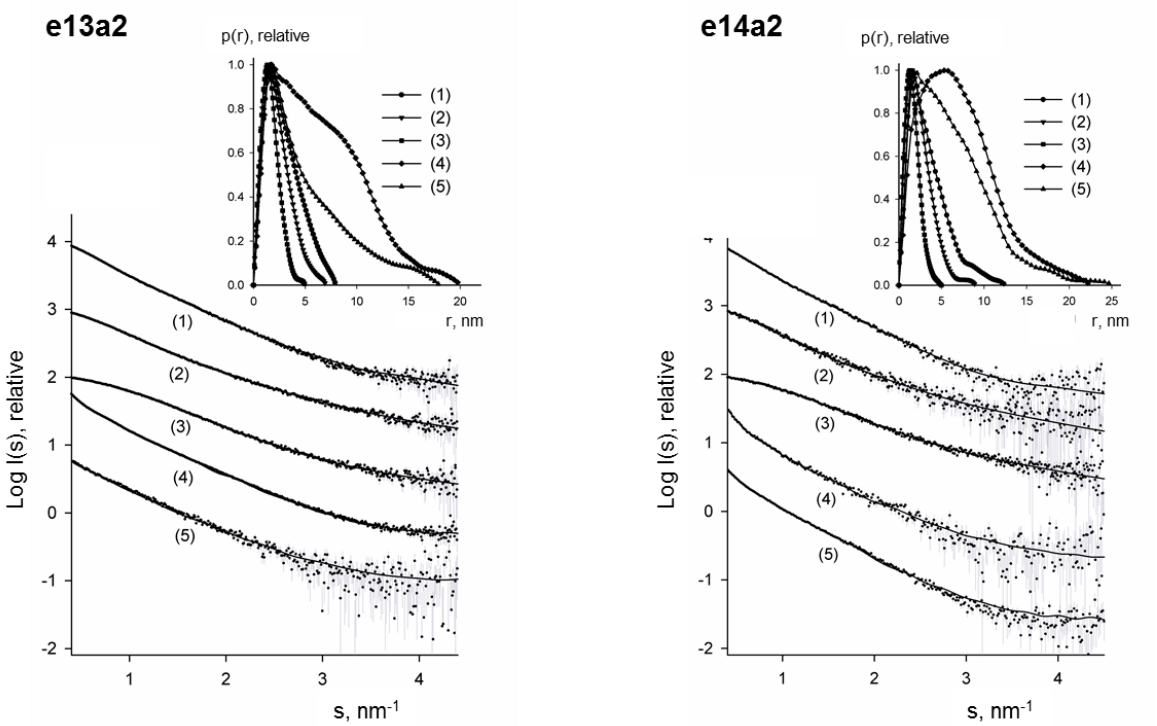


Figure 4.2. SAXS experimental scattering data (dots) and scattering calculated from the *ab initio* models (continuous line) for e13a2 (left) and 314a2 (right). The inserts contain distance distribution functions $P(r)$. 1 - e13A or e14A; 2 - e13B or e14B; 3 - e13C or e14C; 4 - e13AB or e14AB, 5 - e13ABC or e14ABC.

After the characterization of the hairpin, target and revelator, a fully complementary target was added and the effect of the hybridization was assessed. The addition of the complementary target has no measurable impact in the emission spectrum of the hairpin, e13A or e14A (solid black line vs dashed green line in Figure A4, Appendix). However, the SAXS data for this situation shows a significant structural alteration, based on the near duplication of the R_g and D_{max} parameters

(e13AB: R_g 5.40 nm and D_{max} 20.0 nm; 14AB: R_g 5.85 nm and D_{max} 22.5 nm) in contrast with the hairpin alone. This duplication is correlated with the disruption of a symmetrical hairpin upon hybridization to the target sequence, forming a more extended structure – Figure 4.2 (4) and Table 4.3. The *ab initio* model of this ensemble was determined, being possible to observed in some regions the DNA double helix – Figure 4.3. Since, the hybridization of the target sequence to the hairpin is 'invisible' by fluorescence, SAXS allows for the first time the 'visualization' of the hairpin disruption, reinforcing the specificity of the design sequence.

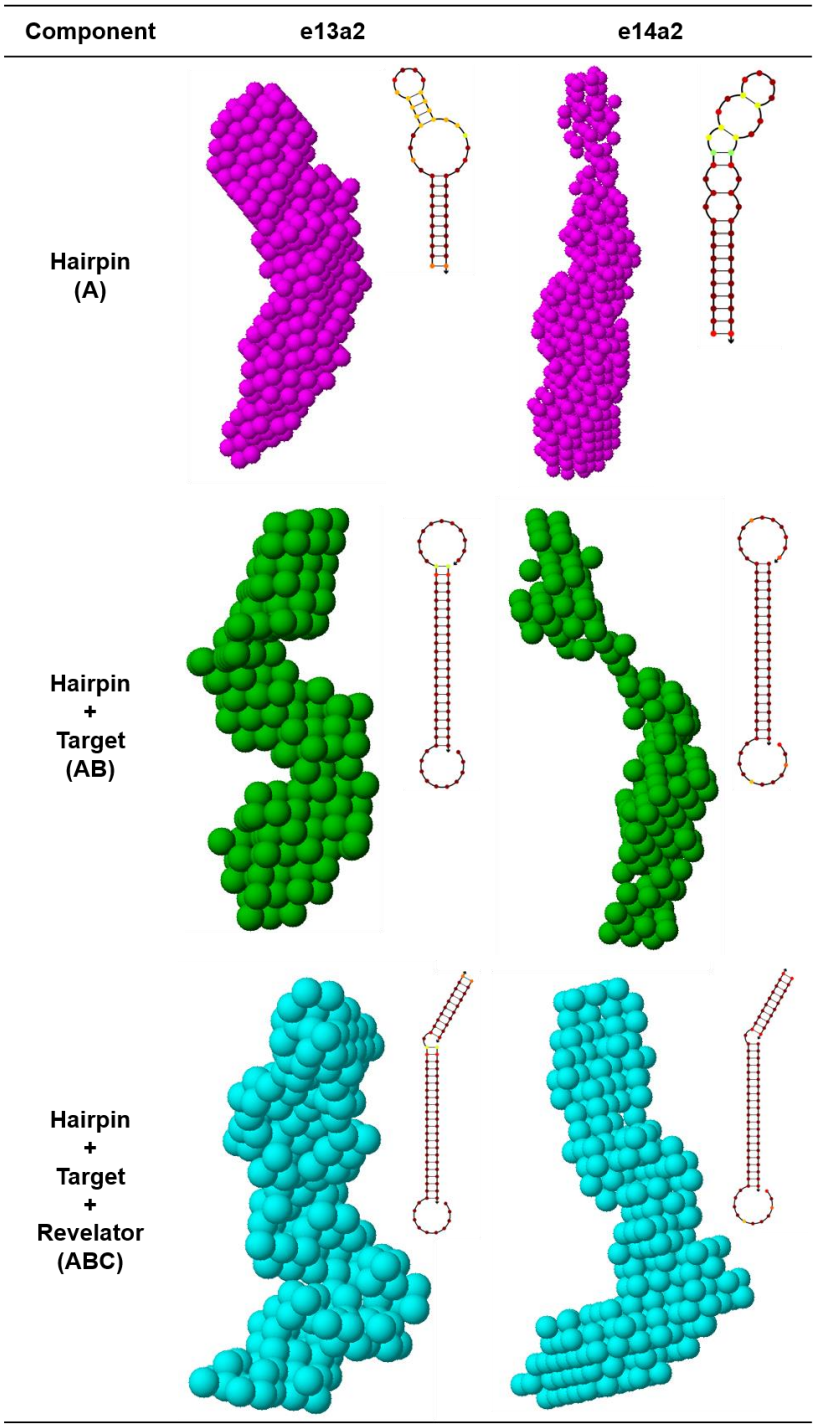


Figure 4.3. *Ab initio* models of the hairpin (magenta), disrupted hairpin after target hybridization (green) and final ensemble (blue). Insets: NUPACK structure prediction of each situation at 283 K.

To test the specificity of the designed system, a partially complementary target was used. Here, the hairpin is in the presence of the corresponding revelator and the partially complementary target (e14A-e13B-e14C and e13A-e14B-e13C) – see the sequences in Table 4.1. For this case, we expect the same results as for the e13A and e14A – considering the same intensity of the donor band as seen in the solid black line vs the dashed blue line in Figure A4, Appendix. However, the R_g and D_{max} values derived from the SAXS curves are higher than the ones obtained for the hairpin alone – Table 4.3. and Figure 4.4. This difference could be derived from the presence of a mixture of free molecules in solution or, on the contrary, to the formation of a 1:1:1 complex. It should be noted that the concentrations and temperature conditions used to perform SAXS and spectral characterization are very different. In the case of SAXS, scattering data is collected at concentrations several folds higher than in an emission spectra assay, which may induce the hybridization of the hairpin with the partial complementary target[306], as simulated by the NUPACK software – Table A2, Appendix. The concentration range used for the SAXS approach is not compatible with fluorescence spectroscopy or with the other techniques used to complement the study, such as microscale thermophoresis or electrophoresis. In this scenario, the collaborators obtained the same spectral signature as the negative control (where the hairpin component, either e13A or e14A, where incubated with a non-complementary target in the presence of the revelator, a situation where the hybridization between strands is not possible) indicating that, in these conditions, the hybridization between the three strands did not occur – dashed blue line in Figure A4, Appendix.

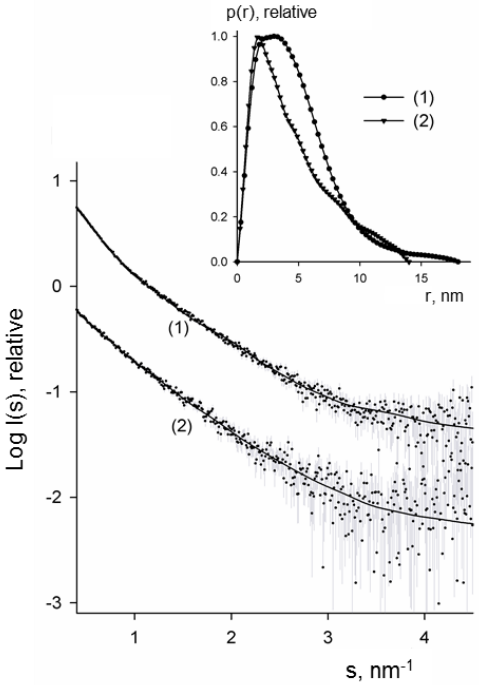


Figure 4.4. SAXS experimental scattering data (dots) and scattering calculated from the *ab initio* models (continuous line) in the presence of the partially complementary sequences. The inset contains distance distribution function $P(r)$. 1 - e13A-e14B-e13C; 2 - e14A-e13B-e14C.

Finally, the three components of the sensor were combined (e13ABC and e14ABC) to form a stable complex. In this case, the presence of a fully complementary target induces a specific FRET signal that can be distinguished from all other scenarios – solid red line Figure A4, Appendix. The R_g and D_{max} values obtained for e14ABC and e13ABC (R_g: 5.21 and 4.56 nm, D_{max}: 20 and 18 nm, respectively) are higher than the ones obtained for the corresponding hairpins alone (e14A and e13A), as expected for an open hairpin hybridized with its target sequence (Table 4.3). Furthermore, these values are slightly lower than the ones observed for e14AB or e13AB, due to the formation of a more compact structure upon the duplex formation at the 3' end of the open hairpin (Table A2, Appendix). This is also observed by the collaborators in the electrophoretic profile (Figure A3, Appendix) using an acrylamide gel, where the complexes e14ABC and e13ABC) migrate beyond the e14AB or e13AB pairs. The *ab initio* bead models obtained for the full biosensor assembly are represented in Figure 4.4. Here, the calculated models reinforce the evidence that e14(e13)ABC forms a more compact structure when compared with e14(e13)AB.

The overall data indicates that the hybridization of the three strands occurs and is sequential since the hybridization of the revelator (e13C or e14C) is dependent on the pre-hybridization of the target (e13B or e14B, respectively). SAXS measurements are concordant with the *in silico* simulations, with the emission spectra and gel electrophoresis, indicating that this technique can be a valuable tool for the characterization of DNA structures.

For detection, the designed hairpins are grafted to the surface of a gold nanoparticle (AuNP). AuNPs have been proposed as effective platforms for bio-sensing, once they can be functionalized with multiple hairpins (in specifying e13a2 and e14a2) and are an easy way to increase the local concentration of the recognition molecule, i.e. hairpin. The AuNP can act as a dark quencher improving the resolution of the fluorescence signals, making the biosensor response clearer.

To study the interaction between the nanoparticle and DNA, SAXS data was collected for the AuNPs (with 14 nm diameter) functionalized with ~70 hairpins (for the variant e13a2) per particle, with the corresponding target and revelator sequences already hybridized - AuABC. The scattering profile derived from a series of dilutions between 10, 5, 2.5 e 1.13 nM in 0.5 × TBE pH 8.3 and 154 mM NaCl were obtained at 277 K – Figure 4.5., applying the same steps described for the different biosensor components in solution – see section 4.2.1.

The R_g and D_{max} parameters obtained (8.0 nm and 28.9 nm, respectively) for this experiment are much larger than those expected for the particles with the mean diameter of 14 nm. These values correspond to the particles with an average diameter of 30 nm. The AuNPs derived from a standardized synthesis (performed by M. Cordeiro) that yields particles of different sizes, with the mean diameter of these AuNP are centred at 14 nm – determined by transmission electron

microscopy. Standard SAXS methods for *ab initio* determination (such as DAMMIF[91]) are not suitable for AuNPs functionalized with DNA data, since the particles and oligo molecules have significantly different scattering densities. The program MONSA[73] from the ATSAS software was used as an alternative, since it allows the fitting simultaneously multiple curves. However, for this one is necessary to assume the monodispersity of the system, which is doubtful due to the polydispersity of AuNPs samples. Using X-rays, the predominant scattering is derived from the metal core, being difficult to detect the biological component. In this context, SANS - Small-Angle Neutron Scattering arise as an alternative. SAXS and SANS are highly complementary methods, whereby the former technique is faster, requires less material and yields usually more precise data. SANS is sensitive to isotopic H/D exchange, which is experimentally used for contrast variation involving measurements in different H_2O/D_2O mixtures, which provides unique information about complex particles[65]. AuNPs functionalized with e13ABC in 100% and 70% of D_2O (Sigma-Aldrich) were prepared and measured at Heinz Maier-Leibnitz (FRM II) at Garching, Germany. The results are under interpretation but in the SANS data for AuABC in 100% D_2O both AuNP and DNA are visible, and for the same sample in 70% D_2O, the AuNP is close to its matching point, so only the DNA contribution is detected. This information can be used to deconvolve the collected SAXS data. The results are promising but need further optimization.

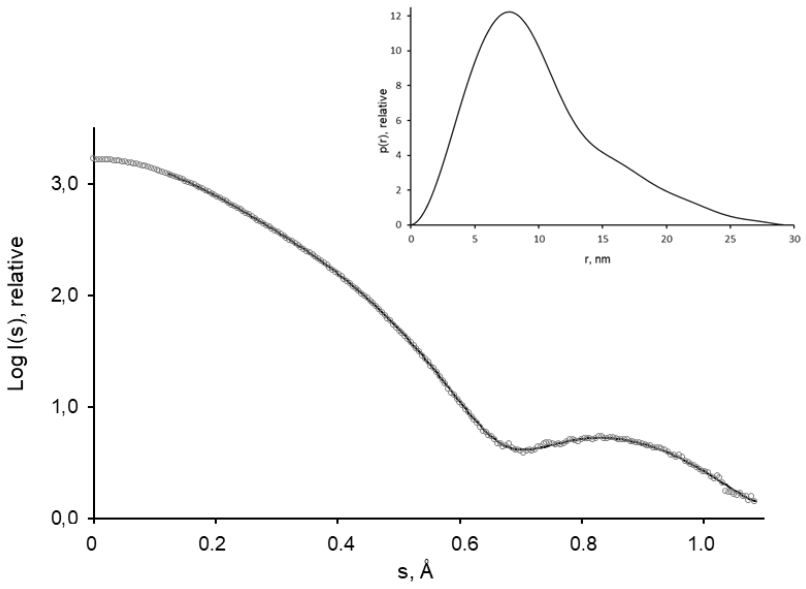


Figure 4.5. SAXS scattering data (points) and GNOM fit (line) for AuNP functionalized with the full biosensor ensemble for e13a2 (hairpin, target and revetator). The data were collected at ESRF (France), beamline BM29. Inset: distance distribution function, $P(r)$.

Chapter 5

Conclusions and future perspectives

5.1. General conclusions

Molybdenum (Mo) and tungsten (W) are trace elements essential to most organisms, from prokaryotes to animals. These metals are found in the active site of important enzymes that catalyze several reactions for carbon, nitrogen and sulfur metabolism. Mo and W are bioavailable as molybdate and tungstate oxyanions, respectively, and are uptaken by the cell through specific ABC-transport systems. Once in the cell, Mo and W are generally coordinated by a unique cofactor, pyranopterin, forming the molybdenum (tungsten) cofactor, Moco (Woco). The Moco is formed by a well-characterized biosynthetic pathway that finishes with the insertion into the corresponding enzymes. Due to the lability of Moco and structural features of the Mo-dependent enzymes, the insertion is, in some cases, mediated by chaperones that also help in protein folding. Structural studies of two proteins involved in the metal selection and transport in *Desulfovibrio alaskensis* G20 (*Da*G20, ModA and TupA) and a molybdoenzyme PaoABC and its chaperone PaoD from *Escherichia coli* were performed using X-ray crystallography and small-angle X-ray scattering (SAXS) as principal techniques.

Several crystal structures of molybdate and tungstate-binding proteins have been solved throughout the years, although, is still not clear how they distinguish between the two similar oxyanions. The analysis of the substrate-binding proteins, ModA and TupA, crystal structures (determined at 1.4 Å and 3.0 Å resolution, respectively) from two independent ABC-transport systems, allowed elucidating which are the residues involved in the ligand binding and important for the selectivity. Despite the differences in the primary structure, the crystal structures of ModA and TupA shared the same architecture as the other substrate-binding proteins, with a butterfly shape and a hinge connecting two lobes. The two lobes are structurally similar, probably enhancing the interaction between the periplasmic component with the dimeric transmembrane component (ModB, TupB). Furthermore, this arrangement also provides the structural flexibility that allows the substrate-binding protein to switch between a loose or compact conformation in the absence or presence of the oxyanion, respectively. The SAXS data clearly reinforce the hypothesis that both proteins adopt a 'Venus Fly-trap' model upon ligand binding, a common feature of the SBP. For TupA, ITC experiments showed the relevance of R118 in the oxyanion binding where three mutants (R118K, R118Q and R118E) lost the ability to bind molybdate. Curiously, for the R118K mutant, the residue substitution increases the selectivity of the protein towards tungstate with high affinity, erasing its ability to bind molybdate. Contrarily to what has been described for TupA wild-type (binds tungstate with a higher affinity than molybdate), ModA binds both oxyanions with very high and similar affinities, being unable to replace one oxyanion for the other in a displacement titration assay.

PaoD is a protein involved in the maturation and insertion of the cofactor into the *Escherichia coli* PaoABC. PaoD has been purified as a dimeric (α_2) protein and could be successfully stabilized and concentrated in the presence of specific ionic liquids (IL) ([C_4mim]Cl and [C_2OHmim]PF_6). To characterize the interaction of the IL moieties with the protein, STD-NMR studies were performed.

The results suggest some degree of directionality and specificity in the interaction IL-protein. The use of ionic liquids for the stabilization and concentration of the chaperone PaoD proved also to be essential for the crystallization assays. The stabilization effect of ILs reduced the precipitation and the size of aggregates during the thawing process, as demonstrated by DLS. They also allowed the protein to be concentrated making it possible to obtain well diffracting crystals. This constitutes an important step towards the structural elucidation of this molecular chaperone.

Despite the dimeric structure of PaoD, the target is the heterotrimeric ($\alpha\beta\gamma$) periplasmic aldehyde oxidoreductase PaoABC. The enzyme was structurally characterized by SAXS and X-ray crystallography. The crystal structure of PaoABC was the first example of a Mo-dependent enzyme containing MCD cofactor in *E. coli* and the first heterotrimer structure of the xanthine oxidase family. During refinement of the crystal structure, the presence of an unanticipated [4Fe-4S] cluster in the FAD subunit (PaoB) was identified. The signal of this cluster was unambiguous after calculation and interpretation of the anomalous maps. This cluster has also been found in the unrelated enzyme 4-HBCR from *Thauera aromatica* and other Moco-containing enzymes with unknown function and structure. These enzymes possess the four cysteine residues required for the binding of the [4Fe-4S] center in approximately 40 residues apart from the insert of the FAD subunit. The similarities suggest a common function for this group of enzymes, that is, the remarkably efficient oxidation of aromatic aldehydes, as observed for PaoABC, although its real physiological function is yet to be determined.

The results obtained and described in this Thesis show the importance of combining different techniques to characterize relevant biological systems. The aims of this Thesis were achieved, with the determination of the tridimensional structure of two substrate-binding proteins and of a molybdoenzyme by X-ray crystallography. SAXS was essential to understand the physiological behavior of the different proteins, showing the flexibility character of ModA and TupA, and reinforcing that PaoABC is one of the few 'monomers' of the xanthine oxidase family.

In addition, SAXS was used to structurally characterize a two-component FRET-based molecular beacon. The duplication of the size of the hairpin upon target recognition and the slight decrease after the hybridization of the revelator, indicated the formation of a 1:1:1 complex formed by the hairpin, target sequence and revelator. SAXS together with other techniques (fluorescence spectroscopy and electrophoretic mobility) shows that the formation of the final complex is sequential and specific, characteristics that are essential for a biosensor.

5.2. Future perspectives

The work presented in this Thesis represents an important step towards the characterization of molybdenum biological pathways with the determination of two crystal structures for a molybdate and tungstate-binding protein and a third structure of a periplasmic aldehyde oxidoreductase from the xanthine oxidase family.

For the molybdate and tungstate-binding proteins, ModA and TupA, is important to determine the crystal structures with the bound oxyanion. The study of two independent ABC transporters for the same molecules in *Da*G20 will give important insights of the selectivity process and coordination to the metal. The high affinity of these SBP for tungstate and molybdate was demonstrated in this Thesis, and especially reinforced by the identification of an arginine important for oxyanion binding in TupA. This mutation increases the selectivity of the protein towards tungstate with high affinity, depleting its ability to bind molybdate. Due to the high affinity and selectivity of the design mutant of TupA, this variation can be considered for potential biotechnological applications. Although still a hypothesis, this mutant would be interesting for further studies, especially when considering the recent reports on tungsten toxicity in humans.

The follow-up of this project is the study of the other constituents of the ABC transporter: the transmembrane domains (TupB, ModB) and the nucleotide-binding domain (TupC, ModC). The work is under development with the financial support of the European Platform iNEXT. The *tupC* and *modC* genes were synthesized by NZYTech (Portugal) and cloned into a pET-system for overexpression in *E. coli*. The first expression tests showed a very large amount of protein produced, but mostly in the insoluble fractions. The expression protocol is currently under optimization, exploring different expression conditions, vectors and bacterial strains. The final goal is to determine the structure of the individual components and complexes (ModABC and TupABC) by SAXS and X-ray crystallography.

There are still many open and important questions directly related to what was explored in this Thesis, particularly regarding the PaoABC and its chaperone PaoD. To answer some of these questions is important to improve the PaoD crystals and solved the 3D structure. A molecular model will allow comparison of the PaoD crystal structure with other chaperones and understand why this protein is a dimer (α_2) when the PaoABC is a heterotrimer ($\alpha\beta\gamma$). Also, the interaction between the two proteins and the cofactor delivering process need to be elucidated, and X-ray crystallography and SAXS are useful techniques for that purpose. Experimental evidences of the PaoD interaction with the protein responsible for the MCD formation should be tested and characterized. After suitable crystallization conditions are established, soaking experiments using cofactor precursor can be used to elucidate protein-ligand interaction and delivery. These results can also be combined with STD-NMR measurements, elucidating the structure of the chaperone-Moco complex and enlightening this interaction.

The interaction between PaoD and the protein responsible for the MCD production, MocA (CTP:molybdopterin cytidylyltransferase), remains unclear. For that, crystallization trials are being performed to solve the MocA crystallographic structure. In the future, SAXS will be used to characterize their interaction in solution.

Advances have also been made regarding the determination of the chaperone FdsD solution structure by NMR. The role of this small protein is unclear, and determining its 3D structure will help to unravel its function. Titration experiments between the FdsD and the FdsC are planned to characterize the interaction between the two proteins involved in the cofactor maturation and understand how they cooperated to produce an active *R. capsulatus* FDH.

E. coli PaoABC is an important enzyme due to its potential biotechnological applications. Contrarily to what has been stated for other molybdoenzymes, the active site of PaoABC is very exposed to the surface. This might be one of the reasons why it has been so challenging to capture a substrate (e.g. cinnamaldehyde) or a ligand (e.g. DFF) bound to its active site. Many conditions and crystals have been tested but, so far, no complex was obtained. The ability to analyze where these compounds bind within the protein active site would be of great interest and crucial for understanding the reaction mechanism. Recently, the spotlights are pointed to PaoABC because of its applicability in the field of biosensors, revealing the urgency to clarify how the reactions occur. The physiological role of the [4Fe-4S] needs also to be addressed and for that, a PaoB-R118A mutant was already expressed, purified and crystallized. The crystals obtained in the same conditions of the wild-type protein diffracted up to 1.7 Å and are under analysis. This residue may facilitate a possible electron transfer route between the [4Fe-4S] cluster and the FAD.

Chapter 6

References

1. Bortels, H. Molybdenum as a catalyst in the biological fixation of nitrogen. *Arch. Mikrobiol.* **1,** 333–342 (1930).
2. Renzo, E. C. De *et al.* Identification of the xanthine oxidase factor as molybdenum. *Arch Biochem Biophys* **45,** 247–253 (1953).
3. Richert, D. A. & Westerfeld, W. W. Isolation and identification of the xanthine oxidase factor as molybdenum. *J Biol Chem* **203,** 915–923 (1953).
4. Ljungdahl, L. G. & Andreesen, J. R. Tungsten, a component of active formate dehydrogenase from Clostridium thermoaceticum. *FEBS Lett.* **54,** 279–282 (1975).
5. Andreesen, J. R. & Ljungdahl, L. G. Formate dehydrogenase of Clostridium thermoaceticum: Incorporation of selenium-75, and the effects of selenite, molybdate, and tungstate on the enzyme. *J. Bacteriol.* **116,** 867–873 (1973).
6. Hille, R. Molybdenum and tungsten in biology. *Trends Biochem. Sci.* **27,** 360–7 (2002).
7. Hille, R., Schulzke, C. & Kirk, M. *Molybdenum and tungsten enzymes: Biochemistry*. *RSC Metallobiology Series* (Royal Society of Chemistry, 2017).
8. Zhang, Y., Rump, S. & Gladyshev, V. N. Comparative genomics and evolution of molybdenum utilization. *Coord. Chem. Rev.* **255,** 1206–1217 (2011).
9. Stiefel, E. I. *The biogeochemistry of molybdenum and tungsten*. *Metal ions in biological systems* (2002).
10. Mendel, R. R. Molybdenum: biological activity and metabolism. *Dalt. Trans.* **21,** 3404–3409 (2005).
11. Laukel, M., Chistoserdova, L., Lidstrom, M. E. & Vorholt, J. A. The tungsten-containing formate dehydrogenase from Methylobacterium extorquens AM1: Purification and properties. *Eur J Biochem* **270,** 325–333 (2003).
12. Mendel, R. R. Biology of the molybdenum cofactor. *J. Exp. Bot.* **58,** 2289–96 (2007).
13. Otrelo-Cardoso, A. R., Nair, R. R., Correia, M. A. S., Rivas, M. G. & Santos-Silva, T. TupA: A tungstate binding protein in the periplasm of Desulfovibrio alaskensis G20. *Int. J. Mol. Sci.* **15,** 11783–98 (2014).
14. Mendel, R. R. & Bittner, F. Cell biology of molybdenum. *Biochim. Biophys. Acta* **1763,** 621–35 (2006).
15. Romão, M. J. Molybdenum and tungsten enzymes: a crystallographic and mechanistic overview. *Dalt. Trans.* 4053–68 (2009).
16. Rajagopalan, K. V. & Jonhson, J. L. The pterin molybdenum cofactors. *J. Biol. Chem.* **267,** 10199–10202 (1992).
17. Hille, R. *Molybdenum enzymes containing the pyranopterin cofactor: An overview*. *Met Ions Biol Syst.* **39,** (2002).
18. Mendel, R. R. & Hänsch, R. Molybdoenzymes and molybdenum cofactor in plants. *J. Exp. Bot.* **53,** 1689–1698 (2002).
19. Kendrew, J. C. *et al.* A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature* **181,** 662–666 (1958).
20. Perutz, M. F. *et al.* Structure of hæmoglobin: a three-dimensional Fourier synthesis at 5.5 Åresolution, obtained by X-ray analysis. *Nature* **185,** 416–422 (1960).
21. Watson, J. D. & Crick, F. H. C. Molecular structure of nucleic acids. A structure for deoxyribose nucleic acid. *Nature* **171,** 737–738 (1953).
22. Shi, Y. A glimpse of structural biology through X-ray crystallography. *Cell* **159,** 995–1014 (2014).
23. Berman, H. M. *et al.* The protein data bank. *Nucleic Acids Res.* **28,** 235–242 (2000).
24. Blundell, T. L. Celebrating structural biology. *Nat. Struct. Mol. Biol.* **18,** 1304–1316 (2011).
25. Wlodawer, A., Minor, W., Dauter, Z., Jaskolski, M. & Physics, B. Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. *FEBS J* **275,** 1–21 (2008).
26. Rose, J. P. & Wang, B. C. SAD phasing: History, current impact and future opportunities. *Arch. Biochem. Biophys.* **602,** 80–94 (2016).
27. Wlodawer, A., Minor, W., Dauter, Z. & Jaskolski, M. Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure

determination. *FEBS J.* **280,** 5705–5736 (2013).

28. Gale, R. *Crystallography made crystal clear - A guide for users of macromolecular models*. (Academic Press, 2006).
29. Ericsson, U. B., Hallberg, B. M., DeTitta, G. T., Dekker, N. & Nordlund, P. Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal. Biochem.* **357,** 289–298 (2006).
30. Proteau, A., Shi, R. & Cygler, M. Application of dynamic light scattering in protein crystallization. *Curr. Protoc. Protein Sci.* **17,** (2010).
31. Rupp, B. *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology*. (Garland Science, 2010).
32. Krauss, I. R., Merlino, A., Vergara, A. & Sica, F. An overview of biological macromolecule crystallization. *Int. J. Mol. Sci.* **14,** 11643–11691 (2013).
33. Kirkwood, J., Hargreaves, D., O'keefe, S. & Wilson, J. Analysis of crystallization data in the Protein Data Bank. *Acta Crystallogr F* **71,** 1228–1234 (2015).
34. Kantardjieff, K. A. & Rupp, B. Protein isoelectric point as a predictor for increased crystallization screening efficiency. *Bioinformatics* **20,** 2162–2168 (2004).
35. Dessau, M. A. & Modis, Y. Protein crystallization for X-ray crystallography. *JoVE* **9,** 1–6 (2011).
36. Chayen, N. E. & Saridakis, E. Protein crystallization: from purified protein to diffraction-quality crystal. *Nat. Methods* **5,** 147–153 (2008).
37. Gavira, J. A. Current trends in protein crystallization. *Arch. Biochem. Biophys.* **602,** 3–11 (2015).
38. McPherson, A. Introduction to protein crystallization. *Methods* **34,** 254–265 (2004).
39. Asherie, N. Protein crystallization and phase diagrams. *Methods* **34,** 266–72 (2004).
40. Malkin, A. J., Kuznetsov, Y. G., Land, T. A., Deyoreo, J. J. & McPherson, A. Mechanisms of growth for protein and virus crystals. *Nat. Struct. Biol.* **2,** 956–959 (1995).
41. Chayen, N. E. Methods for separating nucleation and growth in protein crystallisation. *Prog. Biophys. Mol. Biol.* **88,** 329–337 (2005).
42. Garciá-Ruiz, J. M. Counterdiffusion methods for macromolecular crystallization. *Methods Enzymol.* **368,** 130–154 (2003).
43. McPherson, A. & Gavira, J. A. Introduction to protein crystallization. *Acta Crystallogr F* **70,** 2–20 (2014).
44. McPherson, A. & DeLucas, L. J. Microgravity protein crystallization. *Microgravity* **1,** 1–20 (2015).
45. Carvalho, A. L., Trincão, J. & Romão, M. J. in *Methods in molecular biology* (ed. Roque, A. C. A.) **572,** (Humana Press, 2009).
46. McPherson, A., Nguyen, C., Cudney, R. & Larson, S. B. The role of small molecule additives and chemical modification in protein crystallization. *Cryst. Growth Des.* **11,** 1469–1474 (2011).
47. Dauter, Z. & Jaskolski, M. How to read (and understand) Volume A of International Tables for Crystallography: An introduction for nonspecialists. *J. Appl. Crystallogr.* **43,** 1150–1171 (2010).
48. Matthews, B. W. Solvent content of protein crystals. *J. Mol. Biol.* **33,** 491–497 (1968).
49. Carvalho, A. L., Trincão, J. & Romão, M. J. *Ligand-Macromolecular Interactions in Drug Discovery, Methods and Protocols*. *Methods in Molecular Biology* **572,** (Humana Press, 2010).
50. Gorrec, F. Protein crystallization screens developed at the MRC Laboratory of Molecular Biology. *Drug Discov. Today* **21,** 819–825 (2016).
51. Lamb, A. L., Kappock, T. J. & Silvaggi, N. R. You are lost without a map: Navigating the sea of protein structures. *Biochim. Biophys. Acta - Proteins Proteomics* **1854,** 258–268 (2015).
52. Bragg, W. H. & Bragg, W. L. The reflections of X-rays by crystals. *Proc. R. Soc. London. Ser. A, Contain. Pap. a Math. Phys. Character.* **88,** 428–438 (1913).
53. Feiten, M. C., Di Luccio, M., Santos, K. F., de Oliveira, D. & Oliveira, J. V. X-ray

crystallography as a tool to determine three-dimensional structures of commercial enzymes subjected to treatment in pressurized fluids. *Appl. Biochem. Biotechnol.* (2016).

54. Hendrickson, W. A. Anomalous Diffraction in Crystallographic Phase Evaluation. *Q Rev Biophys* **47,** 49–93 (2014).
55. Dauter, Z. & Jaskol. Crystal pathologies in macromolecular crystallography. *Postepy Biochem.* **62,** 401–407 (2016).
56. Taylor, G. L. Introduction to phasing. *Acta Crystallogr D* **66,** 325–338 (2010).
57. Blow, D. M. & Rossmann, M. G. The single isomorphous replacement method. *Acta Crystallogr.* **14,** 1195–1202 (1961).
58. Pusey, M. L. *et al.* Life in the fast lane for protein crystallization and X-ray crystallography. *Prog. Biophys. Mol. Biol.* **88,** 359–386 (2005).
59. Tickle, I., Sharff, A., Vinkovic, M., Yon, J. & Jhoti, H. High-throughput protein crystallography and drug discovery. *Chem. Soc. Rev.* **33,** 558–65 (2004).
60. Taylor, G. The phase problem. *Acta Crystallogr D* **59,** 1881–1890 (2003).
61. Lemke, C. T., Smith, G. D. & Howell, P. L. S-SAD, Se-SAD and S/Se-SIRAS using Cu Ka radiation: why wait for synchrotron time? *Acta Crystallogr D* **D58,** 2096–2101 (2002).
62. Xu, H. *et al.* Away from the edge II: in-house Se-SAS phasing with chromium radiation. *Acta Crystallogr D* **D61,** 960–966 (2005).
63. Rose, J. P., Wang, B. & Weiss, M. S. Native SAD is maturing. *IUCrJ* **2,** 431–440 (2015).
64. Touw, W. G., Joosten, R. P. & Vriend, G. New biological insights from better structure models. *J. Mol. Biol.* **428,** 1375–1393 (2016).
65. Svergun, D. I. Small-angle X-ray and neutron scattering as a tool for structural systems biology. *Biol. Chem.* **391,** 737–743 (2010).
66. Guinier, A. La diffraction des rayons X aux tres petits angles: application a l'etude de phenomenes ultramicroscopiques. *Ann Phys* **12,** 161–237 (1939).
67. Guinier, A. & Fournet, G. *Small-angle scattering of X-rays*. *J Polym Sci* **1,** (John Wiley & Sons, Inc., 1955).
68. Blanchet, C. E. & Svergun, D. I. Small-angle X-ray scattering on biological macromolecules and nanocomposites in solution. *Annu. Rev. Phys. Chem.* **64,** 37–54 (2013).
69. Porod, G. *General theory*. *Small Angle X-ray Scattering* (Academic Press, 1982).
70. Koch, M. H. J., Vachette, P. & Svergun, D. I. Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. *Q. Rev. Biophys.* **36,** 147–227 (2003).
71. Trewhella, J. Small-angle scattering and 3D structure interpretation. *Curr. Opin. Struct. Biol.* **40,** 1–7 (2016).
72. Lipfert, J. & Doniach, S. Small-angle X-ray scattering from RNA, proteins, and protein complexes. *Annu. Rev. Biophys. Biomol. Struct.* **36,** 307–327 (2007).
73. Svergun, D. I. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76,** 2879–2886 (1999).
74. Mertens, H. D. T. & Svergun, D. I. Structural characterization of proteins and complexes using small-angle X-ray solution scattering. *J. Struct. Biol.* **172,** 128–41 (2010).
75. Volkov, V. V. & Svergun, D. I. DI. Uniqueness of ab initio shape determination in small-angle scattering. *Int. Union Crystallogr.* **36,** 860–864 (2003).
76. Kikhney, A. G. & Svergun, D. I. A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Lett.* **589,** 2570–2577 (2015).
77. Lee, C. *et al.* Structural basis for inhibition of the replication licensing factor Cdt1 by geminin. *Nature* **430,** 913–917 (2004).
78. De Marco, V. *et al.* Quaternary structure of the human Cdt1-Geminin complex regulates DNA replication licensing. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 19807–19812 (2009).
79. Pollack, L. & Doniach, S. *Time-resolved X-ray scattering and RNA folding*. *Methods in enzymology* **469,** (Elsevier Inc., 2009).
80. Putnam, C. D., Hammel, M., Hura, G. L. & Tainer, J. A. X-ray solution scattering (SAXS)

combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.* **40,** 191–285 (2007).

81. Franke, D., Kikhney, A. G. & Svergun, D. I. Automated acquisition and analysis of small angle X-ray scattering data. *Nucl. Instruments Methods Phys. Res. A* **689,** 52–59 (2012).
82. Grant, T. D. *et al.* The accurate assessment of small-angle X-ray scattering data. *Acta Crystallogr D* **71,** 45–56 (2015).
83. Jacques, D. A. & Trewhella, J. Small-angle scattering for structural biology - Expanding the frontier while avoiding the pitfalls. *Protein Sci.* **19,** 642–657 (2010).
84. Petoukhov, M. V., Konarev, P. V., Kikhney, A. G. & Svergun, D. I. ATSAS 2.1 – towards automated and web- supported small-angle scattering data analysis. *J. Appl. Crystallogr.* **40,** 223–228 (2007).
85. Petoukhov, M. V. *et al.* New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **45,** 342–350 (2012).
86. Petoukhov, M. V & Svergun, D. I. Applications of small-angle X-ray scattering to biomacromolecular solutions. *Int. J. Biochem. Cell Biol.* **45,** 429–37 (2013).
87. Mylonas, E. & Svergun, D. Accuracy of molecular mass determination of proteins in solution by small-angle X-ray scattering. *J. Appl. Crystallogr.* **40,** s245–s249 (2007).
88. Svergun, D. I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25,** 495–503 (1992).
89. Putnam, C. D. Guinier peak analysis for visual and automated inspection of small-Angle X-ray scattering data. *J. Appl. Crystallogr.* **49,** 1412–1419 (2016).
90. Chacón, P., Morán, F., Díaz, J. F., Pantos, E. & Andreu, J. M. Low-resolution structures of proteins in solution retrieved from X-ray scattering with a genetic algorithm. *Biophys. J.* **74,** 2760–2775 (1998).
91. Franke, D. & Svergun, D. I. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Cryst.* **43,** 342–346 (2009).
92. Svergun, D. I., Petoukhov, M. V & Koch, M. H. Determination of domain structure of proteins from X-ray solution scattering. *Biophys. J.* **80,** 2946–2953 (2001).
93. Perry, J. J. P. & Tainer, J. A. Developing advanced X-ray scattering methods combined with crystallography and computation. *Methods* **59,** 363–371 (2013).
94. Svergun, D., Barberato, C. & Koch, M. H. J. CRYSOL – A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **28,** 768–773 (1995).
95. Petoukhov, M. V & Svergun, D. I. Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys. J.* **89,** 1237–1250 (2005).
96. Otrelo-Cardoso, A. R. *et al.* Structural data on the periplasmic aldehyde oxidoreductase PaoABC from Escherichia coli: SAXS and preliminary X-ray crystallography analysis. *Int. J. Mol. Sci.* **15,** 2223–36 (2014).
97. Rees, D. C., Johnson, E. & Lewinson, O. ABC transporters: the power to change. *Nat. Rev. Mol. Cell Biol.* **10,** 218–227 (2009).
98. Rego, A. C. & Oliveira, C. R. de. *Membranas biológicas e transporte*. *Bioquímica: Organização Molecular da Vida* (Lidel, 2008).
99. Blattner, F. R., Plunkett, G. I., Bloch, C. A. & Perna, N. T. The complete genome sequence of Escherichia coli K-12. *Science (80-.).* **277,** 1453–1462 (1997).
100. Busch, W. & Saier, M. H. The transporter classification (TC) system, 2002. *Crit. Rev. Biochem. Mol. Biol.* **37,** 287–337 (2002).
101. Locher, K. P. Structure and mechanism of ATP-binding cassette transporters. *Phil. Trans. R. Soc. B* **364,** 239–245 (2009).
102. Dawson, R. J. P. & Locher, K. P. Structure of a bacterial multidrug ABC transporter. *Nature* **443,** 180–185 (2006).
103. Linton, K. J. Structure and function of ABC transporters. *Physiology* **22,** 122–30 (2007).
104. ter Beek, J., Guskov, A. & Slotboom, D. J. Structural diversity of ABC transporters. *J. Gen. Physiol.* **143,** 419–35 (2014).

105. Higgins, C. F. ABC transporters: from microorganisms to man. *Annu. Rev. Cell Biol.* **8,** 67–113 (1992).
106. Biemans-Oldehinkel, E., Doeven, M. K. & Poolman, B. ABC transporter architecture and regulatory roles of accessory domains. *FEBS Lett.* **580,** 1023–1035 (2006).
107. El-Awady, R. *et al.* The role of eukaryotic and prokaryotic ABC transporter family in failure of chemotherapy. *Front Pharmacol* **7,** 1–15 (2017).
108. Lewinson, O. & Livnat-Levanon, N. Mechanism of action of ABC importers: conservation, divergence, and physiological adaptations. *J. Mol. Biol.* **429,** 606–619 (2017).
109. Durmort, C. & Brown, J. S. *Streptococcus pneumoniae: Lipoproteins and ABC transporters*. *Streptococcus Pneumoniae* (Elsevier Inc., 2015).
110. Linton, K. J. & Higgins, C. F. The Escherichia coli ATP-binding cassette (ABC) proteins. *Mol Microbiol* **28,** 5–13 (1998).
111. Hopfner, K.-P. Architectures and mechanisms of ATP binding cassette proteins. *Biopolymers* **105,** 492–504 (2016).
112. Ames, G. F. Bacterial periplasmic transport systems: structure, mechanism, and evolution. *Annu. Rev. Biochem.* **55,** 397–425 (1986).
113. Wilkens, S. Structure and mechanism of ABC transporters. *F1000Prime Rep.* **7,** 1–9 (2015).
114. Locher, K. P., Lee, A. T. & Rees, D. C. The E. coli BtuCD structure: A framework for ABC transporter architecture and mechanism. *Science (80-.).* **296,** 1091–1098 (2002).
115. Hollenstein, K., Frei, D. C. & Locher, K. P. Structure of an ABC transporter in complex with its binding protein. *Nature* **446,** 213–6 (2007).
116. Hvorup, R. N. *et al.* Asymmetry in the structure of the ABC transporter–binding protein complex BtuCD-BtuF. *Science (80-.).* **317,** 1387–1390 (2007).
117. Xu, K. *et al.* Crystal structure of a folate energy-coupling factor transporter from Lactobacillus brevis. *Nature* **497,** 268–272 (2013).
118. Beis, K. Structural basis for the mechanism of ABC transporters. *Biochem Soc Trans* **43,** 889–893 (2015).
119. Quiocho, F. A. & Ledvina, P. S. Atomic structure and specificity of bacterial periplasmic receptors for active transport and chemotaxis: Variation of common themes. *Mol. Microbiol.* **20,** 17–25 (1996).
120. Mulligan, C., Fischer, M. & Thomas, G. H. Tripartite ATP-independent periplasmic (TRAP) transporters in bacteria and archaea. *FEMS Microbiol Rev* **35,** 68–86 (2011).
121. Winnen, B., Hvorup, R. N. & Jr, M. H. S. The tripartite tricarboxylate transporter (TTT) family. *Res. Microbiol. 154* **154,** 457–465 (2003).
122. Scheepers, G. H., Lycklama, J. A. & Poolman, B. An updated structural classification of substrate-binding proteins. *FEBS Lett.* **590,** 4393–4401 (2016).
123. Berntsson, R. P. A., Smits, S. H. J., Schmitt, L., Slotboom, D. J. & Poolman, B. A structural classification of substrate-binding proteins. *FEBS Lett.* **584,** 2606–2617 (2010).
124. Shuman, H. A. Active transport of maltose in Escherichia coli K12. *J. Biol. Chem.* **257,** 5455–5461 (1982).
125. Davidson, A. L., Dassa, E., Orelle, C. & Chen, J. Structure, function, and evolution of bacterial ATP-binding cassette systems. *Microbiol. Mol. Biol. Rev.* **72,** 317–64 (2008).
126. Brass, J. M., Boos, W. & Hengge, R. Reconstitution of maltose transport in malB mutants of Escherichia coli through calcium-induced disruptions of the outer membrane. *J. Bacteriol.* **146,** 10–17 (1981).
127. Heide, T. Van Der & Poolman, B. ABC transporters: one, two or four extracytoplasmic substrate-binding sites? *EMBO Rep.* **3,** 938–943 (2002).
128. Zhang, P., Wang, J. & Shi, Y. Structure and mechanism of the S component of a bacterial ECF transporter. *Nature* **468,** 717–720 (2010).
129. Rodionov, D. A. *et al.* A novel class of modular transporters for vitamins in prokaryotes. *J. Bacteriol.* **191,** 42–51 (2009).
130. Quiocho, F.A., Phillips, G.N., Spurlino, J.C. and Rodseth, L. E. Crystallographic data of an L-arabinose-binding protein from Escherichia coli. *J Mol Biol* **86,** 491–493 (1974).

131. Marinelli, F. *et al.* Evidence for an allosteric mechanism of substrate release from membrane-transporter accessory binding proteins. *Proc. Natl. Acad. Sci. U. S. A.* **108,** (2011).

132. Fukami-Kobayashi, K., Tateno, Y. & Nishikawa, K. Domain dislocation: a change of core structure in periplasmic binding proteins in their evolutionary history. *J. Mol. Biol.* **286,** 279–290 (1999).

133. Lee, Y., Deka, R. K., Michael, V., Radolf, J. D. & Hasemann, C. A. Treponema pallidum TroA is a periplasmic zinc-binding protein with a helical backbone. *Nature* **6,** 628–633 (1999).

134. Fischer, M., Zhang, Q. Y., Hubbard, R. E. & Thomas, G. H. Caught in a TRAP: substrate-binding proteins in secondary transport. *Trends Microbiol.* **18,** 471–478 (2010).

135. Quiocho, F. A., Spurlino, J. C. & Rodseth, L. E. Extensive features of tight oligosaccharide binding revealed in high-resolution structures of the maltodextrin transport / chemosensory receptor. *Structure* **5,** 997–1015 (1997).

136. Chen, X. *et al.* Structural identification of a bacterial quorum-sensing signal containing boron. *Nature* **415,** 545–549 (2002).

137. Berntsson, R. P. *et al.* The structural basis for peptide selection by the transport receptor OppA. *EMBO J.* **28,** 1–9 (2009).

138. Levdikov, V. M. *et al.* The structure of the oligopeptide-binding protein, AppA, from Bacillus subtilis in complex with a nonapeptide. *J Mol Biol* **345,** 879–892 (2005).

139. Lawson, D. M., Williams, C. E., Mitchenall, L. A. & Pau, R. N. Ligand size is a major determinant of specificity in periplasmic oxyanion-binding proteins: the 1.2 A resolution crystal structure of Azotobacter vinelandii ModA. *Structure* **6,** 1529–1539 (1998).

140. Soriano, E. V *et al.* Structural similarities between thiamin-binding protein and thiaminase-I suggest a common ancestor. *Biochemistry* **47,** 1346–1357 (2008).

141. Kuhlmann, S. I., Scheltinga, A. C. T. van, Bienert, R., Kunte, H.-J. & Ziegler, C. 1.55 Å Structure of the ectoine binding protein TeaA of the osmoregulated TRAP-transporter TeaABC from Halomonas elongata. *Biochemistry* **47,** 9475–9485 (2008).

142. Rucktooa, P. *et al.* Crystal structures of two Bordetella pertussis periplasmic receptors contribute to defining a novel Pyroglutamic Acid Binding DctP subfamily. *J Mol Biol* **370,** 93–106 (2007).

143. Schiefner, A. *et al.* Cation- Interactions as determinants for binding of the compatible solutes glycine betaine and proline betaine by the periplasmic ligand-binding protein ProX from Escherichia coli. *J. Biol. Chem.* **279,** 5588–5596 (2004).

144. Oh, B.-H. *et al.* The bacterial periplasmic histidine-binding protein. *J. Biol. Chem.* **269,** 4135–4143 (1994).

145. Smits, S. H. J. *et al.* The compatible-solute-binding protein OpuAC from Bacillus subtilis: Ligand binding, site-directed mutagenesis, and crystallographic Studies. *J. Bacteriol.* **190,** 5663–5671 (2008).

146. Wolters, J. C. *et al.* Ligand binding and crystal structures of the substrate- binding domain of the ABC transporter OpuA. *PLoS One* **5,** e10361–e10361 (2010).

147. Hollenstein, K., Dawson, R. J. & Locher, K. P. Structure and mechanism of ABC transporter proteins. *Curr. Opin. Struct. Biol.* **17,** 412–418 (2007).

148. Jardetzky, O. Simple allosteric model for membrane pumps. *Nature* **211,** 969–970 (1966).

149. Locher, K. P. Mechanistic diversity in ATP-binding cassette (ABC) transporters. *Nat. Struct. Mol. Biol.* **23,** 487–493 (2016).

150. Yu, Y. *et al.* Planar substrate-binding site dictates the specificity of ECF-type nickel/cobalt transporters. *Cell Res.* **24,** 267–77 (2014).

151. Karpowich, N. K., Song, J. M. & Cocco, N. ATP binding drives substrate capture in an ECF transporter by a release and catch mechanism. *Nat. Struct. Mol. Biol.* **22,** 565–571 (2015).

152. Vetter, I. R. & Wittinghofer, a. Nucleoside triphosphate-binding proteins: different scaffolds to achieve phosphoryl transfer. *Q. Rev. Biophys.* **32,** 1–56 (1999).

153. Patzlaff, J. S., Van der Heide, T. & Poolman, B. The ATP/substrate stoichiometry of the ATP-binding cassette (ABC) transporter OpuA. *J. Biol. Chem.* **278,** 29546–29551 (2003).

154. Oswald, C., Holland, I. B. & Schmitt, L. The motor domains of ABC-transporters: What can structures tell us? *Naunyn. Schmiedebergs. Arch. Pharmacol.* **372,** 385–399 (2006).

155. Schwarz, G., Mendel, R. R. & Ribbe, M. W. Molybdenum cofactors, enzymes and pathways. *Nature* **460,** 839–47 (2009).

156. Hagen, W. R. Cellular uptake of molybdenum and tungsten. *Coord. Chem. Rev.* **255,** 1117–1128 (2011).

157. Rech, S., Wolin, C. & Gunsalus, R. P. Properties of the periplasmic ModA molybdate-binding protein of Escherichia coli. *J. Biol. Chem.* **271,** 2557–62 (1996).

158. Bevers, L. E., Hagedoorn, P.-L., Krijger, G. C. & Hagen, W. R. Tungsten transport protein A (WtpA) in Pyrococcus furiosus: the first member of a new class of tungstate and molybdate transporters. *J. Bacteriol.* **188,** 6498–505 (2006).

159. Anderson, L. A. *et al.* Characterisation of the molybdenum-responsive ModE regulatory protein and its binding to the promoter region of the modABCD (molybdenum transport) operon of *Escherichia coli*. *Eur J Biochem* **246,** 119–126 (1997).

160. Rech, S., Deppenmeier, U. & Gunsalus, R. P. Regulation of the molybdate transport operon, modABCD, of *Escherichia coli* in response to molybdate availability. *J Bacteriol* **177,** 1023–1029 (1995).

161. Schwarz, G., Hagedoorn, P. L. & Fischer, K. *Molybdate and tungstate: Uptake, homeostasis, cofactors, and enzymes*. *Microbiology of Heavy Metals* (2007).

162. Coelho, C. *et al.* Structural insights into xenobiotic and inhibitor binding to human aldehyde oxidase. *Nat. Chem. Biol.* **11,** 779–83 (2015).

163. Gates, A. J. *et al.* Properties of the periplasmic nitrate reductases from *Paracoccus pantotrophus* and *Escherichia coli* after growth in tungsten-supplemented media. *FEMS Microbiol Lett* **220,** 261–269 (2003).

164. May, H. D., Patel, P. S. & Ferry, J. G. Effect of molybdenum and tungsten on synthesis and composition of formate dehydrogenase in *Methanobacterium formicicum*. *J Bacteriol* **170,** 3384–3389 (1988).

165. Pollock, V. V, Conover, R. C., Johnson, M. K. & Barber, M. J. Bacterial expression of the molybdenum domain of assimilatory nitrate reductase: production of both the functional molybdenum-containing domain and the nonfunctional tungsten analog. *Arch Biochem Biophys* **403,** 237–248 (2002).

166. Makdessi, K., Andreesen, J. R. & Pich, A. Tungstate uptake by a highly specific ABC transporter in *Eubacterium acidaminophilum*. *J Biol Chem* **276,** 24557–24564 (2001).

167. Imperial, J., Hadi, M. & Amy, N. K. Molybdate binding by ModA, the periplasmic component of the Escherichia coli mod molybdate transport system. *Biochim. Biophys. Acta* **1370,** 337–46 (1998).

168. Bevers, L. E., Schwarz, G. & Hagen, W. R. A molecular basis for tungstate selectivity in prokaryotic ABC transport systems. *J. Bacteriol.* **193,** 4999–5001 (2011).

169. Hollenstein, K. *et al.* Distorted octahedral coordination of tungstate in a subfamily of specific binding proteins. *J. Biol. Inorg. Chem.* **14,** 663–72 (2009).

170. Hu, Y., Rech, S., Gunsalus, R. P. & Rees, D. C. Crystal structure of the molybdate binding protein ModA. *Nature Structural & Molecular Biology* **4,** 703–707 (1997).

171. Andreesen, J. R. & Makdessi, K. Tungsten, the surprisingly positively acting heavy metal element for prokaryotes. *Ann. N. Y. Acad. Sci.* **1125,** 215–229 (2008).

172. Smart, J. P., Cliff, M. J. & Kelly, D. J. A role for tungsten in the biology of Campylobacter jejuni: Tungstate stimulates formate dehydrogenase activity and is transported via an ultra-high affinity ABC system distinct from the molybdate transporter. *Mol. Microbiol.* **74,** 742–757 (2009).

173. Tejada-Jiménez, M., Chamizo-Ampudia, A., Galván, A., Fernández, E. & Llamas, Á. Molybdenum metabolism in plants. *Metallomics* **5,** 1191–203 (2013).

174. Tejada-Jiménez, M., Llamas, A., Sanz-Luque, E., Galván, A. & Fernández, E. A high-affinity molybdate transporter in eukaryotes. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 20126–30 (2007).

175. Feio, M. J. *et al.* Desulfovibrio alaskensis sp. nov., a sulphate-reducing bacterium from a soured oil reservoir. *Int. J. Syst. Evol. Microbiol.* **54,** 1747–1752 (2004).

176. Hauser, L. J. *et al.* Complete genome sequence and updated annotation of Desulfovibrio alaskensis G20. *J. Bacteriol.* **193,** 4268–4269 (2011).
177. Enning, D. & Garrelfs, J. Corrosion of iron by sulfate-reducing bacteria: New views of an old problem. *Appl. Environ. Microbiol.* **80,** 1226–1236 (2014).
178. Greene, E. A., Brunelle, V., Jenneman, G. E. & Voordouw, G. Synergistic inhibition of microbial sulfide production by combinations of the metabolic inhibitor nitrite and biocides. *Appl. Environ. Microbiol.* **72,** 7897–7901 (2006).
179. Nemati, M., Mazutinec, J. T., Jenneman, E. G. & Voordouw, G. Control of biogenic H2S production with nitrite and molybdate. *J. Ind. Microbiol. Biotechnol.* **26,** 350–355 (2001).
180. Mota, C. S. *et al.* Effects of molybdate and tungstate on expression levels and biochemical characteristics of formate dehydrogenases produced by Desulfovibrio alaskensis NCIMB 13491. *J. Bacteriol.* **193,** 2917–23 (2011).
181. Nair, R. R. *et al.* Changes in metabolic pathways of Desulfovibrio alaskensis G20 cells induced by molybdate excess. *JBIC J. Biol. Inorg. Chem.* **20,** 311–322 (2014).
182. Jancarik, J. & Kim, S. H. Sparse matrix sampling: a screening method for crystallization of proteins. *J Appl Crystallogr* **24,** 409–411 (1991).
183. Kabsch, W. Xds. *Acta Crystallogr D* **66,** 125–32 (2010).
184. Evans, P. R. An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr D* **67,** 282–92 (2011).
185. Number 4, C. C. P. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D* **50,** 760–63 (1994).
186. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).
187. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D* **53,** 240–55 (1997).
188. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D* **60 (Pt 12,** 2126–32 (2004).
189. Joosten, R. P., Long, F., Murshudov, G. N. & Perrakis, A. The PDB_REDO server for macromolecular structure model optimization. *IUCrJ* **1,** 213–220 (2014).
190. Terwilliger, T. C. *et al.* Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr D* **64,** 61–69 (2007).
191. Incardona, M. F. *et al.* EDNA: A framework for plugin-based applications applied to X-ray experiment online data analysis. *J. Synchrotron Radiat.* **16,** 872–879 (2009).
192. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36,** 1277–1282 (2003).
193. Panjkovich, A. & Svergun, D. I. Deciphering conformational transitions of proteins by small angle X-ray scattering and normal mode analysis. *Phys. Chem. Chem. Phys.* **18,** 5707–5719 (2016).
194. Valentini, E., Kikhney, A. G., Previtali, G., Jeffries, C. M. & Svergun, D. I. SASBDB, a repository for biological small-angle scattering data. *Nucleic Acids Res.* **43,** D357–D363 (2015).
195. Mehtab, S. *et al.* Interaction of vanadium(IV) with human serum apo-transferrin. *J. Inorg. Biochem.* **121,** 187–95 (2013).
196. Velazquez-Campoy, A. & Freire, E. Isothermal titration calorimetry to determine association constants for high-affinity ligands. *Nat. Protoc.* **1,** 186–91 (2006).
197. DeLano, W. The PyMOL Molecular Graphics System. *San Carlos, CA DeLano Sci.* (2002).
198. Mao, B., Pear, M. R. & McCammon, J. A. Hinge-bending in L-Arabinose- binding Protein. *J. Biol. Chem.* **257,** 1131–1133 (1982).
199. De Beer, T. A. P., Berka, K., Thornton, J. M. & Laskowski, R. A. PDBsum additions. *Nucleic Acids Res.* **42,** 292–296 (2014).
200. Ledvina, P. S., Yaot, N., Choudhary, A. & Quiocho, F. A. Negative electrostatic surface potential of protein sites specific for anionic ligands. *Biochemistry* **93,** 6786–6791 (1996).
201. Balan, A. *et al.* Crystallographic structure and substrate-binding interactions of the

molybdate-binding protein of the phytopathogen Xanthomonas axonopodis pv. citri. *Biochim. Biophys. Acta* **1784,** 393–9 (2008).

202. Chan, S. *et al.* Apo and ligand-bound structures of ModA from the archaeon Methanosarcina acetivorans. *Acta Crystallogr F* **66,** 242–50 (2010).
203. Dereeper, A. *et al.* Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* **36,** 465–469 (2008).
204. Franke, D., Jeffries, C. M. & Svergun, D. I. Correlation Map, a goodness-of-fit test for one-dimensional X-ray scattering spectra. *Nat. Methods* **12,** 419–422 (2015).
205. Kozin, M. B. & Svergun, D. I. Automated matching of high- and low-resolution structural models. *J. Appl. Crystallogr.* **34,** 33–41 (2001).
206. Krainer, G. & Keller, S. Single-experiment displacement assay for quantifying high-affinity binding by isothermal titration calorimetry. *Methods* **76,** 116–123 (2015).
207. Kletzin, A. & Adams, M. W. W. Tungsten in biological systems. *FEMS Microbiol. Rev.* **18,** 5–63 (1996).
208. Johnson, J. L. & Rajagopalan, K. V. Structural and metabolic relationship between the molybdenum cofactor and urothione. *Proc. Natl. Acad. Sci. U. S. A.* **79,** 6856–6860 (1982).
209. Leimkühler, S. Shared function and moonlighting proteins in molybdenum cofactor biosynthesis. *Biol. Chem.* **398,** 1009–1026 (2017).
210. Nason, A. *et al.* In vitro formation of assimilatory reduced nicotinamide adenine dinucleotide phosphate: nitrate reductase from a Neurospora mutant and a component of molybdenum-enzymes. *Proc. Natl. Acad. Sci. U. S. A.* **68,** 3242–3246 (1971).
211. Reschke, S. *et al.* Identification of a bis-molybdopterin intermediate in molybdenum cofactor biosynthesis in Escherichia coli. *J. Biol. Chem.* **288,** 29736–29745 (2013).
212. Neumann, M., Seduk, F., Iobbi-Nivol, C. & Leimkühler, S. Molybdopterin dinucleotide biosynthesis in Escherichia coli: identification of amino acid residues of molybdopterin dinucleotide transferases that determine specificity for binding of guanine or cytosine nucleotides. *J Biol Chem* **286,** 1400–1408 (2011).
213. Wuebbens, M. M. & Rajagopalan, K. V. Structural characterization of a molybdopterin precursor. *J. Biol. Chem.* **268,** 13493–8 (1993).
214. Clinch, K. *et al.* Synthesis of cyclic pyranopterin monophosphate, a biosynthetic intermediate in the molybdenum cofactor pathway. *J. Med. Chem.* **56,** 1730–1738 (2013).
215. Schwarz, G. Molybdenum cofactor and human disease. *Curr. Opin. Chem. Biol.* **31,** 179–187 (2016).
216. Mendel, R. R. & Magalon, A. Biosynthesis and insertion of the molybdenum cofactor. *EcoSal Plus* **6,** (2015).
217. Mendel, R. R. & Schwarz, G. Molybdenum cofactor biosynthesis in plants and humans. *Coord. Chem. Rev.* **255,** 1145–1158 (2011).
218. Schwarz, G. & Mendel, R. R. Molybdenum cofactor biosynthesis and molybdenum enzymes. *Annu. Rev. Plant Biol.* **57,** 623–47 (2006).
219. Dahl, J. *et al.* The sulfur carrier protein TusA has a pleiotropic role in Escherichia coli that also affects molybdenum cofactor. *J. Biol. Chem.* **288,** 5426–5442 (2013).
220. Mendel, R. R. The molybdenum cofactor. *J. Biol. Chem.* **288,** 13165–13172 (2013).
221. Nichols, J. & Rajagopalan, K. V. Escherichia coli MoeA and MogA: Function in metal incorporation step of molybdenum cofactor biosynthesis. *J. Biol. Chem.* **277,** 24995–5000 (2002).
222. Lake, M. W., Temple, C. a, Rajagopalan, K. V & Schindelin, H. The crystal structure of the Escherichia coli MobA protein provides insight into molybdopterin guanine dinucleotide biosynthesis. *J. Biol. Chem.* **275,** 40211–7 (2000).
223. Neumann, M., Mittelstädt, G., Seduk, F., Iobbi-Nivol, C. & Leimkühler, S. MocA is a specific cytidylyltransferase involved in molybdopterin cytosine dinucleotide biosynthesis in Escherichia coli. *J Biol Chem* **284,** 21891–21898 (2009).
224. Hille, R., Hall, J. & Basu, P. The mononuclear molybdenum enzymes. *Chem. Rev.* **114,** 3963–4038 (2014).
225. Feng, C., Tollin, G. & Enemark, J. H. Sulfite oxidizing enzymes. *Biochim. Biophys. Acta*

1774, 527–539 (2007).

226. Kisker, C. *et al.* Molecular basis of sulfite oxidase deficiency from the structure of sulfite oxidase. *Cell* **91,** 973–983 (1997).
227. Schrader, N. *et al.* The crystal structure of plant sulfite oxidase provides insights into sulfite oxidation in plants and animals. *Structure* **11,** 1251–1263 (2003).
228. Kappler, U. & Bailey, S. Molecular basis of intramolecular electron transfer in sulfite-oxidizing enzymes is revealed by high resolution strcuture of a heterodimeric complex of the catalytic molybdopterin subunit and a c-type cytochrome subunit. *J. Biol. Chem.* **280,** 24999–25007 (2005).
229. Hänsch, R. *et al.* Plant Sulfite Oxidase as Novel Producer of H2O2. *J. Biol. Chem.* **281,** 6884–6888 (2006).
230. Llamas, A., Chamizo-Ampudia, A., Tejada-Jimenez, M., Galvan, A. & Fernandez, E. The molybdenum cofactor enzyme mARC: Moonlighting or promiscuous enzyme? *BioFactors* 1–9 (2017).
231. Schindelin, H., Kisker, C., Hilton, J., Rajagopalan, K. V & Rees, D. C. Crystal structure of DMSO reductase: redox-linked changes in molybdopterin coordination. *Science (80-.).* **272,** 1615–1621 (1996).
232. Schneider, F. *et al.* Crystal structure of dimethyl sulfoxide reductase from Rhodobacter capsulatus at 1.88 A resolution. *J. Mol. Biol.* **263,** 53–69 (1996).
233. Magalon, A., Fedor, J. G., Walburger, A. & Weiner, J. H. Molybdenum enzymes in bacteria and their maturation. *Coord. Chem. Rev.* **255,** 1159–1178 (2011).
234. Brondino, C. D., Romão, M. J., Moura, I. & Moura, J. J. G. Molybdenum and tungsten enzymes: the xanthine oxidase family. *Curr Opin Chem Biol* **10,** 109–114 (2006).
235. Pacher, P., Nivorozhkin, A. & Szabó, C. Therapeutic effects of xanthine oxidase inhibitors: Renaissance half a century after the discovery of Allopurinol. *Pharm. Rev.* **58,** 87–114 (2006).
236. Terao, M. *et al.* Structure and function of mammalian aldehyde oxidases. *Arch. Toxicol.* **90,** 753–780 (2016).
237. Enroth, C., Eger, B. T., Okamoto, K., Nishino, T. & Pai, E. F. Crystal structures of bovine milk xanthine dehydrogenase and xanthine oxidase: structure-based mechanism of conversion. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 10723–8 (2000).
238. Unciuleac, M., Warkentin, E., Page, C. C., Boll, M. & Ermler, U. Structure of a xanthine oxidase-related 4-hydroxybenzoyl-CoA reductase with an additional [4Fe-4S] cluster and an inverted electron flow. *Structure* **12,** 2249–56 (2004).
239. Schumann, S. *et al.* The mechanism of assembly and cofactor insertion into Rhodobacter capsulatus xanthine dehydrogenase. *J. Biol. Chem.* **283,** 16602–11 (2008).
240. Romão, M. J. *et al.* Crystal structure of the xanthine oxidase-related aldehyde oxido-reductase from D. gigas. *Science (80-.).* **270,** 1170–6 (1995).
241. Marangon, J. *et al.* Kinetic and structural studies of aldehyde oxidoreductase from Desulfovibrio gigas reveal a dithiolene-based chemistry for enzyme activation and inhibition by H2O2. *PLoS One* **8,** (2013).
242. Boll, M. *et al.* Redox centers of 4-hydroxybenzoyl-CoA reductase, a member of the xanthine oxidase family of molybdenum-containing enzymes. *J. Biol. Chem.* **276,** 47853–62 (2001).
243. Neumann, M. *et al.* A periplasmic aldehyde oxidoreductase represents the first molybdopterin cytosine dinucleotide cofactor containing molybdo-flavoenzyme from Escherichia coli. *FEBS J* **276,** 2762–2774 (2009).
244. Truglio, J. J. *et al.* Crystal structures of the active and alloxanthine-inhibited forms of xanthine dehydrogenase from Rhodobacter capsulatus. *Structure* **10,** 115–125 (2002).
245. Dobbek, H., Gremer, L., Meyer, O. & Huber, R. Crystal structure and mechanism of CO dehydrogenase, a molybdo iron-sulfur flavoprotein containing S-selanylcysteine. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 8884–9 (1999).
246. Hanzelmann, P. *et al.* The effect of intracellular molybdenum in Hydrogenophaga pseudoflava on the crystallographic structure of the seleno-molybdo-iron-sulfur flavoenzyme carbon monoxide dehydrogenase. *J. Mol. Biol.* **301,** 1221–1235 (2000).

247. Bonin, I. *et al.* Active site geometry and substrate recognition of the molybdenum hydroxylase quinoline 2-oxidoreductase. *Structure* **12,** 1425–35 (2004).

248. Loschi, L. *et al.* Structural and biochemical identification of a novel bacterial oxidoreductase. *J. Biol. Chem.* **279,** 50391–400 (2004).

249. Gennaris, A. *et al.* Repairing oxidized proteins in the bacterial envelope using respiratory chain electrons. *Nature* **528,** 409–412 (2015).

250. Badalyan, A. *et al.* Analysis of the interaction of the molybdenum hydroxylase PaoABC from Escherichia coli with positively and negatively charged metal complexes. *Electrochem. commun.* **37,** 5–7 (2013).

251. Badalyan, A., Neumann-Schaal, M., Leimkühler, S. & Wollenberger, U. A biosensor for aromatic aldehydes comprising the mediator dependent PaoABC-aldehyde oxidoreductase. *Electroanalysis* **25,** 101–108 (2013).

252. Badalyan, A. *et al.* Electrical wiring of the aldehyde oxidoreductase PaoABC with a polymer containing osmium redox centers: Biosensors for benzaldehyde and GABA. *Biosensors* **4,** 403–421 (2014).

253. Leimkühler, S. & Klipp, W. Role of XDHC in molybdenum cofactor insertion into xanthine dehydrogenase of Rhodobacter capsulatus. *J Bacteriol* **181,** 2745–2751 (1999).

254. Leimkühler, S., Angermüller, S., Schwarz, G., Mendel, R. R. & Klipp, W. Activity of the molybdopterin-containing xanthine dehydrogenase of Rhodobacter capsulatus can be restored by high molybdenum concentrations in a moeA mutant defective in molybdenum cofactor biosynthesis. *J. Bacteriol.* **181,** 5930–9 (1999).

255. Neumann, M., Stöcklein, W., Walburger, A., Magalon, A. & Leimkühler, S. Identification of a Rhodobacter capsulatus L-cysteine desulfurase that sulfurates the molybdenum cofactor when bound to XdhC and before its insertion into xanthine dehydrogenase. *Biochemistry* **46,** 9586–9595 (2007).

256. Neumann, M., Stöcklein, W. & Leimkühler, S. Transfer of the molybdenum cofactor synthesized by Rhodobacter capsulatus MoeA to XdhC and MobA. *J Biol Chem* **282,** 28493–28500 (2007).

257. Iobbi-Nivol, C. & Leimkühler, S. Molybdenum enzymes , their maturation and molybdenum cofactor biosynthesis in Escherichia coli. *Biochim Biophys Acta* **1827,** 1086–1101 (2012).

258. Otrelo-Cardoso, A. R. *et al.* Biochemical, stabilization and crystallization studies on a molecular chaperone (PaoD) involved in the maturation of molybdoenzymes. *PLoS One* **9,** e87295 (2014).

259. Gasteiger, E. *et al. Protein Identification and Analysis Tools on the ExPASy Server*. *The Proteomics Protocols Handbook* (2005).

260. Viegas, A., Manso, J., Nobrega, F. L. & Cabrita, E. J. Saturation-Transfer Difference (STD) NMR: A simple and fast method for ligand screening and characterization of protein binding. *J Chem Educ* **88,** 990–994 (2011).

261. Meyer, B. & Peters, T. NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors. *Angew Chem Int Ed.* **42,** 864–890 (2003).

262. Yuan, Y., Wen, X., Sanders, D. a R. & Pinto, B. M. Exploring the mechanism of binding of UDP-galactopyranose to UDP-galactopyranose mutase by STD-NMR spectroscopy and molecular modeling. *Biochemistry* **44,** 14080–9 (2005).

263. Leslie, A. G. W. Recent changes to the MOSFLM package for processing film and image plate data. *Jt. CCP4 + ESF-EAMCB Newsl. Protein Crystallogr.* (1992).

264. Evans, P. R. Scaling and assessment of data quality. *Acta Crystallogr D* **62,** 72–82 (2006).

265. Kabsch, W. Evaluation of single-crystal X-ray diffraction data from a position-sensitive detector. *J Appl Crystallogr* **21,** 916–24 (1988).

266. Coelho, C., Trincão, J. & João Romão, M. The use of ionic liquids as crystallization additives allowed to overcome nanodrop scaling up problems: A success case for producing diffraction-quality crystals of a nitrate reductase. *J Cryst Growth* **312,** 714–719 (2010).

267. Böhmer, N., Hartmann, T. & Leimkühler, S. The chaperone FdsC for Rhodobacter capsulatus formate dehydrogenase binds the bis-molybdopterin guanine dinucleotide cofactor. *FEBS Lett.* **588,** 531–7 (2014).

268. Mann, J. P., McCluskey, A. & Atkin, R. Activity and thermal stability of lysozyme in alkylammonium formate ionic liquids—influence of cation modification. *Green Chem* **11,** 785–792 (2009).

269. Byrne, N. & Angell, C. A. Protein unfolding, and the 'tuning in' of reversible intermediate states, in protic ionic liquid media. *J Mol Biol* **378,** 707–114 (2008).

270. Kowacz, M. *et al.* Hofmeister effects of ionic liquids in protein crystallization: Direct and water-mediated interactions. *CrystEngComm* 1–10 (2012).

271. Weingärtner, H. Understanding ionic liquids at the molecular level: facts, problems, and controversies. *Angew. Chem. Int. Ed. Engl.* **47,** 654–70 (2008).

272. Judge, R. A. *et al.* The effect of ionic liquids on protein crystallization and X-ray diffraction resolution. *Cryst. Growth Des.* **9,** 3463–3469 (2009).

273. Kumar, A., Bisht, M. & Venkatesu, P. Biocompatibility of ionic liquids towards protein stability: A comprehensive overview on the current understanding and their implications. *Int. J. Biol. Macromol.* **96,** 611–651 (2017).

274. Ji, Z., Yao, Z. & Liu, M. Saturation transfer difference nuclear magnetic resonance study on the specific binding of ligand to protein. *Anal Biochem* **385,** 380–2 (2009).

275. Zwart, P. H., Grosse-Kunstleve, R. W. & Adams, P. D. Xtriage and Fest: automatic assessment of X-ray data and substructure structure factor estimation. *CCP4 Newsl.* (2005).

276. Powell, H. *et al.* Mosflm 7.0.1 and its new interface - iMosflm 0.5.3. *CCP4 Newsl. Protein Crystallogr.* (2007).

277. Kabsch, W. Automatic indexing of rotation diffraction patterns. *J. Appl. Crystallogr.* **21,** 67–72 (1988).

278. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J Mol Biol* **215,** 403–410 (1990).

279. Pauff, J. M., Cao, H. & Hille, R. Substrate orientation and catalysis at the molybdenum site in xanthine oxidase. Crystal structures in complex with xanthine and lumazine. *J. Biol. Chem.* **284,** 8760–8767 (2009).

280. Cao, H., Hall, J. & Hille, R. X-ray crystal structure of arsenite-inhibited xanthine oxidase: '-sulfifo, ?-oxo double bridge between molybdenum and arsenic in the active site. *J. Am. Chem. Soc.* **133,** 12414–12417 (2011).

281. Cowtan, K. D. & Zhang, K. Y. Density modification for macromolecular phase improvement. *Prog. Biophys. Mol. Biol.* **72,** 245–70 (1999).

282. Laskowski, R. A., MacArthur, M. W., Moss, D. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst* **26,** 283–291 (1993).

283. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D* **66,** 12–21 (2010).

284. Blanchet, C. E. *et al.* Instrumental setup for high-throughput small- and wide-angle solution scattering at the X33 beamline of EMBL Hamburg. *J. Appl. Crystallogr.* **45,** 489–495 (2012).

285. Sticht, H. & Rösch, P. The structure of iron-sulfur proteins. *Prog. Biophys. Mol. Biol.* **70,** 95–136 (1998).

286. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372,** 774--797 (2007).

287. Correia, M. A. S. *et al.* The Escherichia coli periplasmic aldehyde oxidoreductase is an exceptional member of the xanthine oxidase family of molybdoenzymes. *ACS Chem. Biol.* **11,** 2923–2935 (2016).

288. McKenna, S. M. *et al.* Enzyme cascade reactions: synthesis of furandicarboxylic acid (FDCA) and carboxylic acids using oxidases in tandem. *Green Chem.* **44,** Ahead of Print (2015).

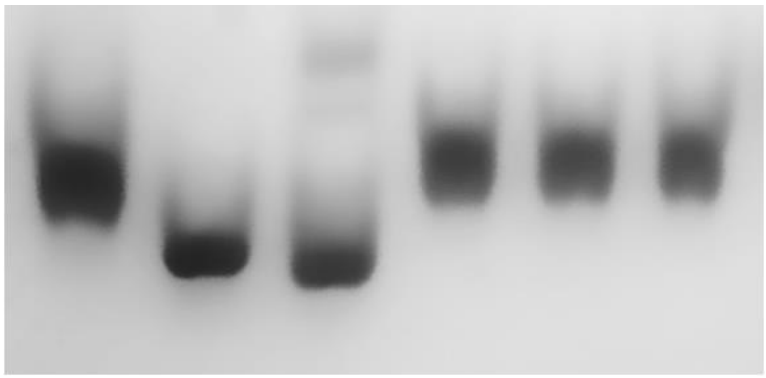
289. Wilner, O. I., Willner, B. & Willner, I. in *Nano-Biotechnology for Biomedical and Diagnostic Research* (eds. Zahavy, E., Ordentlich, A., Yitzhaki, S. & Shafferman, A.) **733,** 97–114 (Springer, 2012).

290. Bruetzel, L. K. *et al.* Conformational changes and flexibility of DNA devices observed by

Small-Angle X-Ray Scattering. *Nano Lett.* **16,** 4871–4879 (2016).

291. Zheng, J. *et al.* Rationally designed molecular beacons for bioanalytical and biomedical applications. *Chem. Soc. Rev.* **44,** 3036–55 (2015).
292. Tyagi, S. & Kramer, F. R. Molecular beacons: probes that fluoresce upon hybridization. *Nat. Biotechnol.* **14,** 303–308 (1996).
293. Singh, J., Kaur, H., Kaushik, A. & Peer, S. A review of antisense therapeutic interventions for molecular biological targets in various diseases. *International Journal of Pharmacology* **7,** 294–315 (2011).
294. Navarro, E., Serrano-Heras, G., Castaño, M. J. & Solera, J. Real-time PCR detection chemistry. *Clin. Chim. Acta* **439,** 231–250 (2015).
295. Thelwell, N., Millington, S., Solinas, A., Booth, J. & Brown, T. Mode of action and application of Scorpion primers to mutation detection. *Nucleic Acids Res.* **28,** 3752–3761 (2000).
296. Nazarenko, I. A., Bhatnagar, S. K. & Hohman, R. J. A closed tube format for amplification and detection of DNA based on energy transfer. *Nucleic Acids Res.* **25,** 2516–2521 (1997).
297. Rosa, J., Conde, J., de la Fuente, J. M., Lima, J. C. & Baptista, P. V. Gold-nanobeacons for real-time monitoring of RNA synthesis. *Biosens bioelectron* **36,** 161–167 (2012).
298. Song, S. *et al.* Gold-nanoparticle-based multicolor nanobeacons for sequence-specific DNA analysis. *Angew. Chemie* **48,** 8670–8674 (2009).
299. Cordeiro, M., Giestas, L., Lima, J. C. & Baptista, P. M. V. BioCode gold-nanobeacon for the detection of fusion transcripts causing chronic myeloid leukemia. *J. Nanobiotechnology* **14,** 38 (2016).
300. Yang, L.-H., Ahn, D. J. & Koo, E. Ultrasensitive FRET-based DNA sensor using PNA/DNA hybridization. *Mater. Sci. Eng. C* **69,** 625–630 (2016).
301. Zhang, C. & Hu, J. Single quantum dot-based nanosensor for multiple DNA detection. *Anal. Chem.* **82,** 1921–7 (2010).
302. Cordeiro, M., Giestas, L., Lima, J. C. & Baptista, P. Coupling an universal primer to SBE combined spectral codification strategy for single nucleotide polymorphism analysis. *J. Biotechnol.* **168,** 90–94 (2013).
303. Koch, W. H. Technology platforms for pharmacogenomic diagnostic assays. *Nat. Rev. Drug Discov.* **3,** 749–761 (2004).
304. Vinhas, R., Cordeiro, M., Pedrosa, P., Fernandes, A. R. & Baptista, P. V. Current trends in molecular diagnostics of chronic myeloid leukemia. *Leuk. Lymphoma* **0,** 1–14 (2016).
305. Zadeh, J. N. *et al.* NUPACK: analysis and design of nucleic acid systems. *J. Comput. Chem.* **31,** 2967–2970 (2010).
306. Markegard, C. B., Gallivan, C. P., Cheng, D. D. & Nguyen, H. D. Effects of concentration and temperature on DNA hybridization by two closely related sequences via large-scale coarse-grained simulations. *J. Phys. Chem. B* **120,** 7795–7806 (2016).

Appendix



1 2 3 4 5 6

Figure A1. Ligand-dependent mobility shift assays for TupA protein (14 µM) in the presence of different oxyanions (10-fold excess). Lane 1: TupA; Lane 2: TupA + MoO_4^{2-}; Lane 3: TupA + WO_4^{2-}; Lane 4: TupA + SO_4^{2-}; Lane 5: TupA + PO_4^{3-}; Lane 6: TupA + ClO_4^-.Gel native TupA

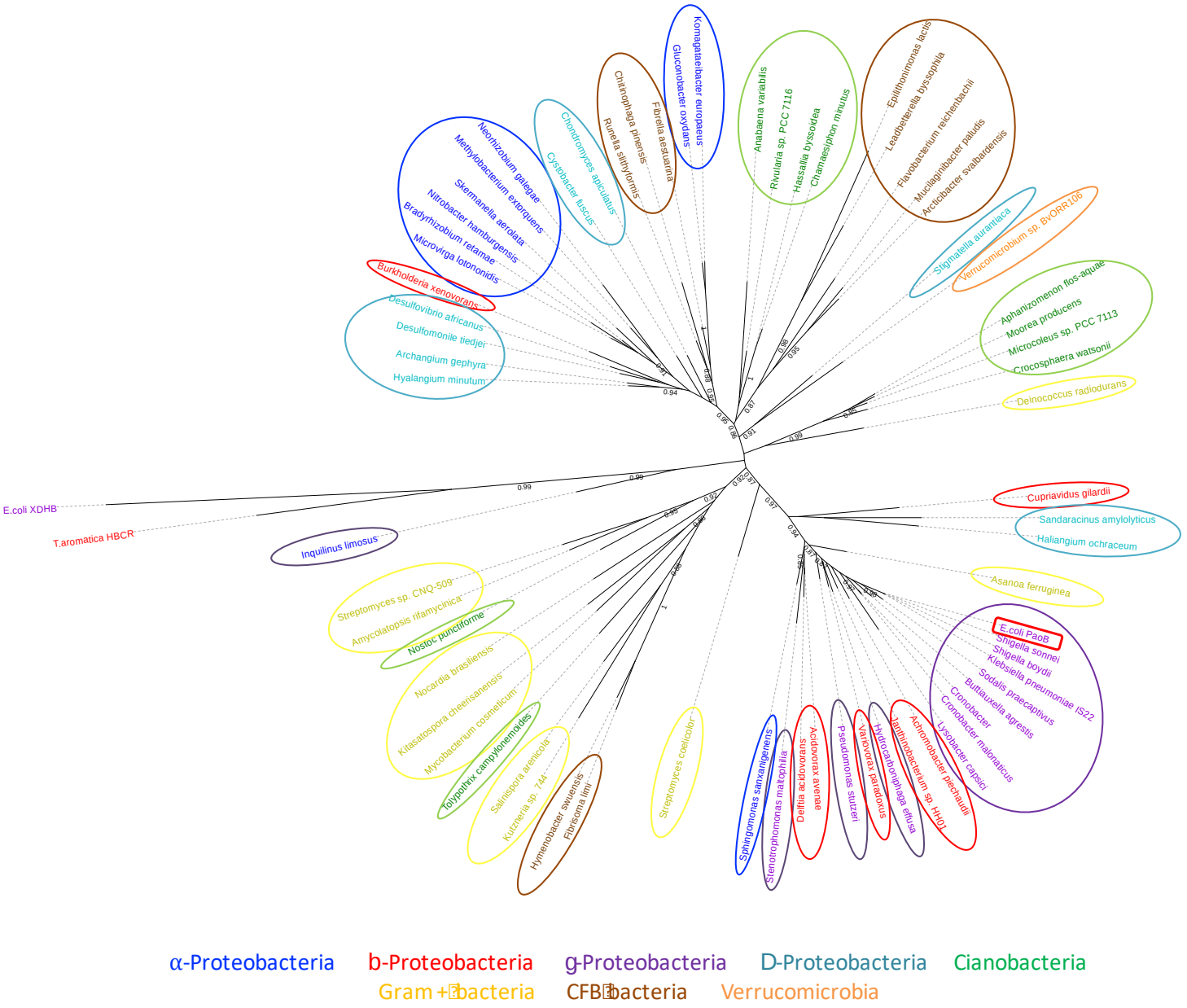
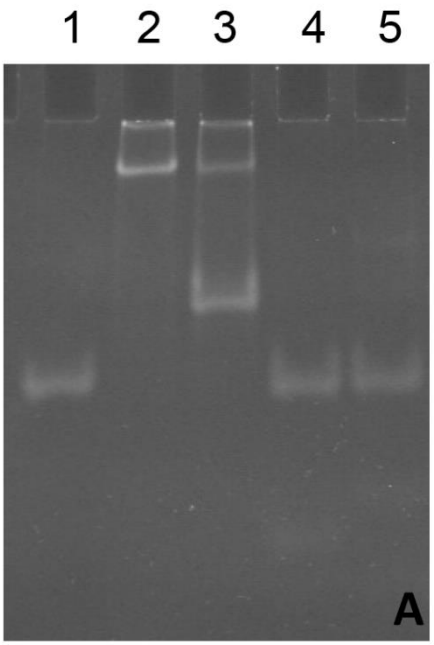
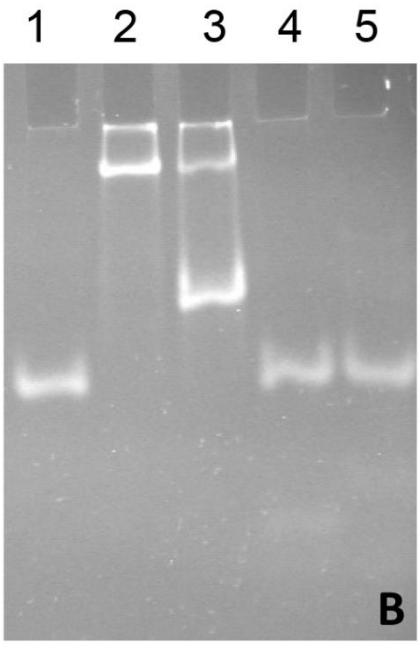


Figure A2. Unrooted dendogram showing distances (represented by branch lengths) for sequences from XO-type enzymes with an additional [4Fe-4S] cluster in FAD subunit. Numbers represent the results of a bootstrap analysis and indicate the number of times out of 100 iterations that these branch points were identified. The FAD subunit of PaoABC (*E. coli* PaoB) is marked with a red rectangle.



	RF
e14A	1
e14AB	0.1809
e14ABC_band1	0.163912
e14ABC_band2	0.689164
e14AYC	1.013774
e14A-e13B-e14C	0.987144



	RF
e13A	1
e13AB	0.177427
e13ABC_band1	0.143056
e13ABC_band2	0.662796
e13AYC	0.974454
e13A-e14B-e13C	1.03019

Figure A3. Acrylamide gel electrophoresis of the tested scenarios. A. Lane 1 – e14A; Lane 2 – e14AB; Lane 3 – e14ABC; Lane 4- e14AYC; Lane 5- e14A-e13B-e14C; B. Lane 1 – e13A; Lane 2 – e13AB; Lane 3 – e13ABC; Lane 4- e13AYC; Lane 5- e13A-e14B-e13C; Running conditions: 20% acrylamide in 1×TBE; 16 hours at 20V, 277 K. RF: relative migration distance.

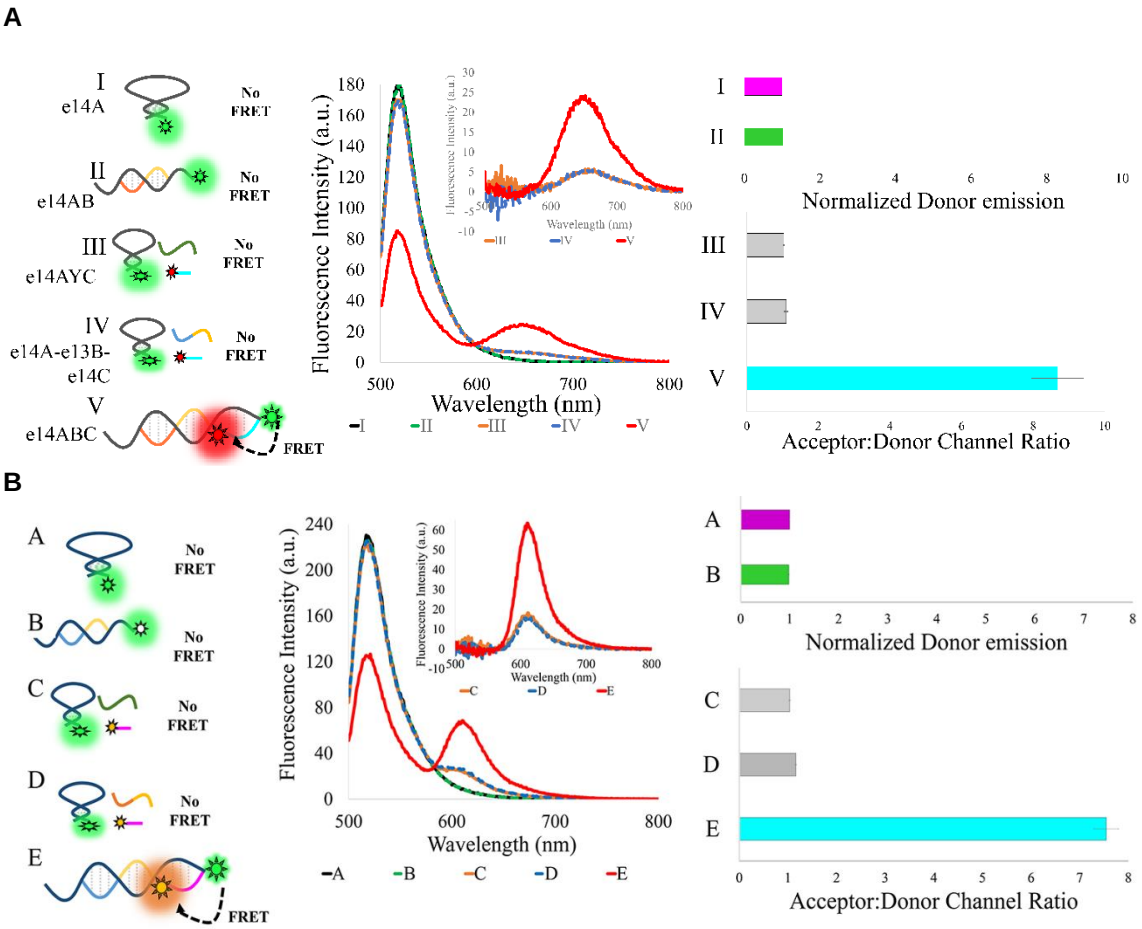


Figure A4. Emission spectra of the two-component molecular beacon (MB) in the tested scenarios. A) MB towards the e14a2 fusion sequence; B) MB towards the e13a2 fusion sequence. Left panel: Schematic representation of the tested scenarios; Middle panel: Raw emission spectra of the tested scenarios; Right Panel: Ratio of acceptor/donor emission for the tested scenarios. I, A – Hairpin alone (solid black line); II, B – Hairpin hybridized to target sequence (dashed green line); III, C – Hairpin in the presence of non-complementary target and revelator (solid orange line); IV, D – Hairpin in the presence of partial complementary target and revelator (dashed blue line); V, E – Hairpin hybridized to target sequence and revelator (solid red line). Inset – emission spectra after donor band subtraction.

Table A1. In-house sparse matrix screen.

1	0.2 M Calcium chloride, 0.1 M Acetate buffer 4.5, 30% 2-Methyl-2,4-pentanediol
2	1 M K/Na tartrate, 0.1 M MES 6.5
3	0.4 M Ammonium phosphate
4	0.1 M Tris-HCl 8.5, 3 M Ammonium sulfate
5	0.2 M Trisodium citrate, 0.1 M HEPES 7.5, 30% 2-Methyl-2,4-pentanediol
6	0.2 M Magnesium chloride, 0.1 M Acetate buffer 4.5, 30% PEG 3350
7	1.2 M Sodium citrate, 0.1 M HEPES buffer 7.5
8	0.2 M Trisodium citrate, 2 M Ammonium sulfate
9	0.2 M Ammonium acetate, 0.1 M Citrate buffer 5.5, 30% PEG 400
10	0.1 M Acetate buffer 4.5, 1.5 M Ammonium phosphate
11	0.2 M Ammonium sulfate, 0.1 M HEPES buffer 7.5, 1.5 M Potassium phosphate/1.5M sodium phosphate
12	0.2 M Trisodium citrate, 0.1 M Tris-HCl 8.5, 20% PEG 400
13	0.2 M Calcium chloride, 0.1 M HEPES buffer 7.5, 25% PEG 3350
14	0.1 M Magnesium chloride, 0.1 M MES buffer 6.5, 30% PEG 8000
15	0.2 M Lithium sulfate, 0.1M Citrate buffer 5.5, 30% PEG 3350
16	1 M Lithium sulfate, 0.1 M Acetate buffer 4.5
17	0.2 M Ammonium phosphate, 0.1 M Tris-HCl 7.5, 30% 2-Methyl-2,4-pentanediol
18	0.2 M Ammonium acetate, 0.1 M Tris-HCl 7.5, 1.5 M Potassium phosphate/1.5M sodium phosphate
19	0.1 M Ammonium sulfate, 0.1 M Citrate buffer 5.5, 30% PEG 8000
20	0.1 M MES buffer 6.5, 30% 2-Methyl-2,4-pentanediol
21	0.2 M Magnesium chloride, 0.1 M HEPES buffer 7.5, 30% PEG 3350
22	0.2 M Sodium acetate, 0.1 M Tris-HCl 8.5, 30% PEG 3350
23	0.1 M Tris-HCl 7.5, 1 M K/Na tartrate
24	0.2 M Calcium chloride, 0.1 M Tris-HCl 8.5
25	0.5 M Ammonium acetate, 0.1 M Citrate buffer 5.5, 30% 2-Methyl-2,4-pentanediol
26	2 M Sodium acetate. 0.1 M MES buffer 6.5
27	0.2 M K/Na tartrate, 0.1 M MES 6.5, 30% PEG 8000
28	1 M K/Na Tartrate, 0.1 M HEPES 7.5
29	0.2 M Ammonium sulfate, 0.1 M Acetate buffer 4.5, 30% PEG 400
30	0.1 M Ammonium sulfate, 0.1 M HEPES buffer 7.5, 20% PEG 3350
31	2 M Ammonium sulfate, 0.1 M MES buffer 6.5
32	0.2 M Sodium chloride, 0.1 M MES 6.5, 30% Ethanol
33	0.2 M Magnesium chloride, 0.1 M HEPES buffer 7.5, 30% Ethanol
34	0.2 M Ammonium acetate, 0.1 M Tris-HCl 8.5, 30% Ethanol
35	0.2 M Calcium chloride, 0.1 M Acetate buffer 4.5, 30% Ethanol
36	0.2 M Sodium acetate, 0.1 M HEPES buffer 7.5, 30% Ethanol
37	0.2 M Magnesium chloride, 0.1 M HEPES 7.5, 30% Isopropanol
38	0.1 M Cacodilate buffer 6.5, 30% 2-Methyl-2,4-pentanediol
39	0.1 M Acetate buffer 4.5, 2 M Sodium formate
40	0.2 M Trisodium citrate, 0.1 M Cacodylate buffer 6.5, 40% isopropanol

41	0.1 M HEPES buffer 7.5, 20 % PEG 400, 10% Isopropanol
42	0.1 M HEPES 7.5, 1M Lithium sulfate
43	0.2 M Lithium sulfate, 0.1 M Tris-HCl 8.5, 30% PEG 3350
44	0.2 M Ammonium sulfate, 0.1 M Cacodylate buffer 6.5, 30% PEG 6000
45	0.1 Acetate buffer 4.5, 1.5 M Sodium acetate
46	0.1 M Trisodium citrate, 1M Ammonium phosphate
47	4 M Sodium formate
48	0.1 M HEPES buffer 7.5, 1.2 M Trisodium citrate
49	0.4 M K/Na tartrate
50	0.2 M Magnesium chloride, 0.1 M Tris-HCl 8.5, 30% PEG 3350
51	0.1 M Cacodylate buffer 6.5, 1.4 M Sodium acetate
52	0.2 M Ammonium acetate, 0.1 M Citrate buffer 5.5, 30% PEG 3350
53	0.2 M Ammonium acetate, 0.1 M Acetate buffer 4.5, 30% PEG 3350
54	0.2 M Calcium chloride, 0.1 M HEPES buffer 7.5, 28% PEG 400
55	0.2 M Ammonium sulfate, 0.1 M Cacodylate buffer 6.5, 30% PEG 8000
56	0.2 M Magnesium acetate, 0.1 M Cacodylate buffer 6.5, 20% PEG 8000
57	0.2 M Ammonium acetate, 0.1 M Tris-HCl 8.5, 30% Isopropanol
58	0.2 M Ammonium sulfate, 0.1 M Acetate buffer 4.5, 25% PEG 3350
59	0.2 M Magnesium acetate, 0.1 M Cacodylate buffer 6.5, 20% 2-Methyl-2,4-pentanediol
60	0.2 M Calcium chloride, 0.1 M Acetate buffer 4.5, 20% Isopropanol
61	0.1 M Imidazole buffer 7.0, 20% Isopropanol
62	0.2 M Trisodium citrate, 0.1 M Cacodylate 6.5, 20% Isopropanol
63	0.2 M Sodium acetate, 0.1 M Cacodylate 6.5, 30% PEG 8000
64	0.2 M Ammonium sulfate, 30% PEG 8000
65	0.2 M Ammonium sulfate. 30% PEG 3350
66	0.1 M HEPES buffer 7.5, 1.6 M K/Na phosphate
67	0.1 M Tris-HCl 8.5, 8% PEG 8000
68	0.1 M Acetate buffer 4.5, 8% PEG 3350
69	0.1 M HEPES buffer 7.5, 2% PEG 400, 2 M Ammonium phosphate
70	0.1 M Citrate buffer 5.5, 20% Isopropanol, 20% PEG 3350
71	0.05 M Potassium phosphate, 20% PEG 8 K
72	30% PEG 8 K
73	0.2 M Magnesium formate
74	0.2 M Zinc acetate, 0.1 M Cacodylate buffer 6.5, 18% PEG 8000
75	0.2 M Calcium acetate, 0.1 M Cacodylate 6.5, 18% PEG 8000
76	0.1 M Acetate buffer 4.5, 2 M Ammonium sulfate
77	0.1 M Tris-HCl 8.5, 2 M Ammonium sulfate
78	1 M Lithium sulfate, 2% PEG 8000
79	0.5 M Lithium sulfate, 15% PEG 8000
80	0.2 M Ammonium acetate, 0.1 M Citrate buffer 5.5, 20% Isopropanol, 20% PEG 3350

Table A2. *In silico* simulations of the designed sequences. The structures were predicted using NUPACK[305] software.

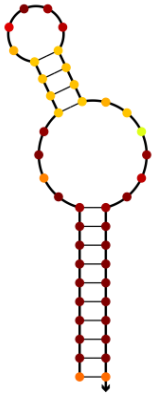
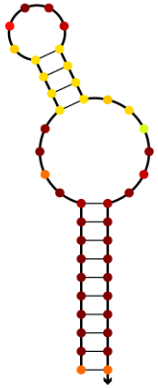
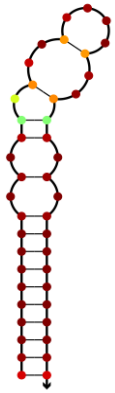
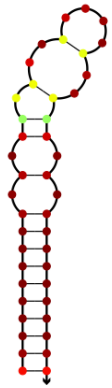
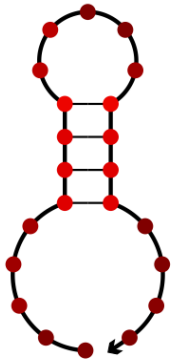
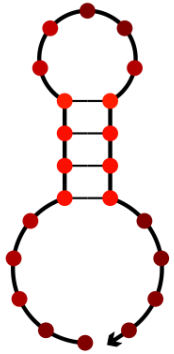
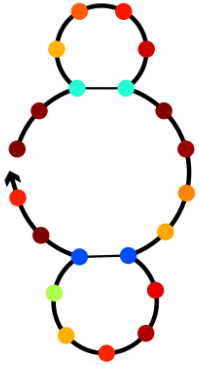
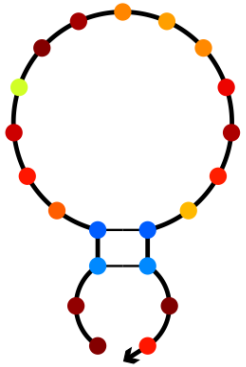
e13a2		e14a2	
Structures at 4 ºC	**Structures at 10 ºC**	**Structures at 4 ºC**	**Structures at 10 ºC**
e13A		e14A	
			
Free energy of secondary structure: -21.77 kcal/mol	Free energy of secondary structure: -19.89 kcal/mol	Free energy of secondary structure: -19.27 kcal/mol	Free energy of secondary structure: -17.08 kcal/mol
e13B		e14B	
			
Free energy of secondary structure: -6.12 kcal/mol	Free energy of secondary structure: -5.44 kcal/mol	Free energy of secondary structure: -0.98 kcal/mol	Free energy of secondary structure: -0.52 kcal/mol

Table A2. *In silico* simulations of the designed sequences (continued). The structures were predicted using NUPACK[305] software.

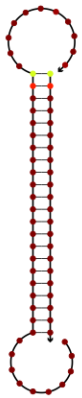
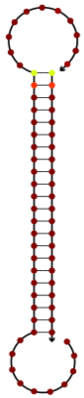
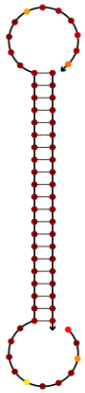
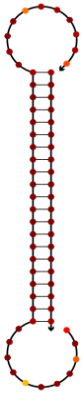
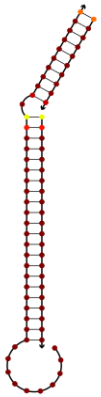
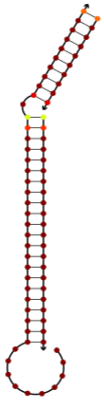
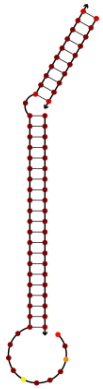
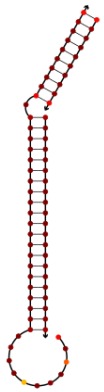
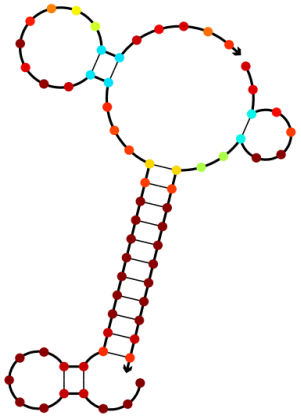
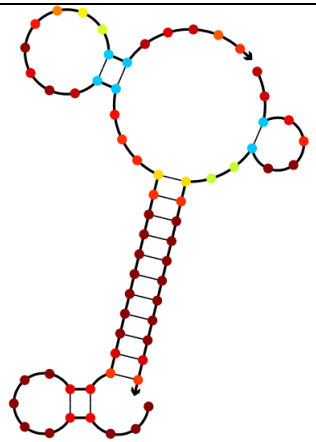
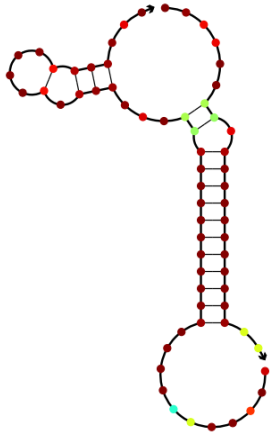
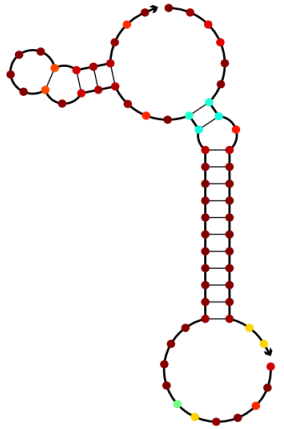
e13		e14	
Structures at 4 ºC	**Structures at 10 ºC**	**Structures at 4 ºC**	**Structures at 10 ºC**
e13AB		e14AB	
			
Free energy of secondary structure: -46.23 kcal/mol	Free energy of secondary structure: -43.47 kcal/mol	Free energy of secondary structure: -43.62 kcal/mol	Free energy of secondary structure: -40.86 kcal/mol
e13ABC		e14ABC	
			
Free energy of secondary structure: -67.70 kcal/mol	Free energy of secondary structure: -63.72 kcal/mol	Free energy of secondary structure: -63.24 kcal/mol	Free energy of secondary structure: -59.32 kcal/mol

Table A2. *In silico* simulations of the designed sequences (continued). The structures were predicted using NUPACK[305] software.

Structures at 4 ºC	Structures at 10 ºC	Structures at 4 ºC	Structures at 10 ºC
e13A-e14B-e13C  Free energy of secondary structure: -26.66 kcal/mol	 Free energy of secondary structure: -24.48 kcal/mol	 Free energy of secondary structure: -28.83 kcal/mol	e14A-e13B-e14C  Free energy of secondary structure: -26.31 kcal/mol